

NAVAL POSTGRADUATE SCHOOL MONTEREY, CALIFORNIA



THESIS

**THE USE OF FINANCIAL SCORING MODELS
FOR THE PREDICTION OF BUSINESS FAILURE:
IMPLICATIONS FOR DEPARTMENT OF
DEFENSE FINANCIAL ANALYSIS**

by

Philip J. Candreva

June 1996

Principal Advisor:

O. D. Moses

Approved for public release; distribution is unlimited

19960805 006

DATA QUALITY ASSURED 1

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 1996	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE THE USE OF FINANCIAL SCORING MODELS FOR THE PREDICTION OF BUSINESS FAILURE: IMPLICATIONS FOR DEPARTMENT OF DEFENSE FINANCIAL ANALYSIS			5. FUNDING NUMBERS	
6. AUTHOR(S) Candрева, Philip J.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The purpose of this thesis is to improve the financial analysis of private sector firms as conducted within the Department of Defense by applying knowledge from the literature related to the use of financial scoring models to predict business failure. First, an original, six-dimensional framework was developed for thoroughly analyzing the literature related to financial scoring models. Second, using the framework, the literature was comprehensively evaluated to assess the state of the art in the use of financial scoring models to predict business failure. Third, the state of financial analysis of private sector firms within DoD was reviewed, both the activities and the methods. Fourth, the literature related to financial analysis within the DoD context was evaluated. Finally, recommendations were made to improve DoD financial analysis based upon the findings in the literature. Those recommendations include a reexamination of the definition of failure and the identification of variables that accurately predict that definition, and the need to construct defensible models.				
14. SUBJECT TERMS Financial Analysis, Business Failure, Financial Scoring Models,			15. NUMBER OF PAGES 116	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

Approved for public release; distribution is unlimited

**THE USE OF FINANCIAL SCORING MODELS
FOR THE PREDICTION OF BUSINESS FAILURE:
IMPLICATIONS FOR DEPARTMENT OF
DEFENSE FINANCIAL ANALYSIS**

Philip J. Candreva
Lieutenant Commander, Supply Corps, United States Navy
B.S., Pennsylvania State University, 1984

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

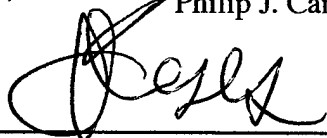
from the

**NAVAL POSTGRADUATE SCHOOL
June 1996**

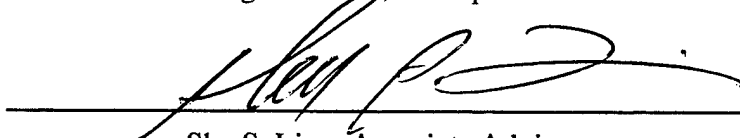
Author: _____


Philip J. Candreva

Approved by: _____



O. Douglas Moses, Principal Advisor



Shu S. Liao, Associate Advisor



Reuben Harris, Chairman
Department of Systems Management

ABSTRACT

The purpose of this thesis is to improve the financial analysis of private sector firms as conducted within the Department of Defense by applying knowledge from the literature related to the use of financial scoring models to predict business failure. First, an original, six-dimensional framework was developed for thoroughly analyzing the literature related to financial scoring models. Second, using the framework, the literature was comprehensively evaluated to assess the state of the art in the use of financial scoring models to predict business failure. Third, the state of financial analysis of private sector firms within DoD was reviewed, both the activities and the methods. Fourth, the literature related to financial analysis within the DoD context was evaluated. Finally, recommendations were made to improve DoD financial analysis based upon the findings in the literature. Those recommendations include a reexamination of the definition of failure and the identification of variables that accurately predict that definition, and the need to construct defensible models.

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	THE RELATIONSHIP BETWEEN THE DEPARTMENT OF DEFENSE AND THE PRIVATE SECTOR.....	1
B.	FINANCIAL SCORING MODELS.....	2
C.	RESEARCH QUESTIONS AND LIMITATIONS.....	3
D.	ORGANIZATION OF STUDY.....	4
II.	FOUNDATION FOR ANALYSIS.....	5
A.	CHARACTERISTICS OF FIRMS THAT FAIL.....	5
B.	WHY STUDY FIRM FAILURE.....	6
C.	FINANCIAL ANALYSIS WITHIN THE DEPARTMENT OF DEFENSE.....	7
D.	SUMMARY.....	8
III.	FRAMEWORK FOR ANALYSIS.....	9
A.	THEORETICAL BASIS.....	9
B.	SAMPLE SELECTION AND DATA COLLECTION.....	10
1.	Conceptual Issues.....	10
2.	Practical Issues.....	12
C.	DEPENDENT VARIABLE.....	13
1.	The Construct Being Investigated.....	13
2.	The Scale of the Output.....	13
D.	INDEPENDENT VARIABLES.....	13
1.	The Information Set.....	14
a.	Qualitative and Quantitative Variables.....	15
b.	Firm Specific or Macroeconomic Variables.....	16
c.	Accounting Data or Independent Analysis.....	16
2.	The Choice of A Specific Measure.....	17
a.	The Construct.....	17
b.	Selection of a Measure.....	17
c.	Data Transformations.....	18
d.	Variable Transformations.....	18
3.	Evaluation Criteria.....	19
a.	Ex Ante Criteria.....	19
b.	Ex Post Criteria.....	19
4.	Summary of Independent Variable Issues.....	20

E.	MODELING TECHNIQUE.....	20
1.	Univariate Discriminant Analysis.....	20
2.	Multivariate Discriminant Analysis.....	20
3.	Conditional Probability Models.....	21
4.	Recursive Partitioning.....	22
5.	Indexing.....	24
6.	Artificial Intelligence.....	24
F.	VALIDATION.....	24
1.	Fit of the Model.....	25
2.	Performance of the Model.....	25
G.	SUMMARY	28
IV.	THE STATE OF THE ART.....	29
A.	THEORETICAL BASIS FOR THE MODEL.....	29
1.	Theory Regarding the Behavior of the Firm.....	29
a.	Cash Flow Theories.....	29
b.	Theories Regarding Failure Events and the Actions of Distressed Firms.....	31
2.	Theories Regarding the Content of Information Sets.....	33
a.	Theories Regarding Financial Ratios.....	33
b.	Theories Derived From the Auditing Literature.....	34
B.	SAMPLE SELECTION AND DATA COLLECTION.....	35
1.	Conceptual Issues.....	35
a.	Firm Size and Industry.....	35
b.	Business and Economic Climate.....	37
c.	Sample Size vs. Relevance.....	38
2.	Practical Issues.....	39
a.	Composition of the Sample.....	39
b.	Availability of Data.....	41
C.	DEPENDENT VARIABLE.....	42
1.	The Construct Being Investigated.....	42
a.	The Definition of Failure.....	42
b.	An Issue of Timing.....	43
2.	The Scale of the Outcome.....	44
D.	INDEPENDENT VARIABLES.....	46
1.	The Information Set.....	46
a.	Qualitative and Quantitative Variables.....	47
b.	Firm Specific or Macroeconomic Variables.....	49
c.	Accounting Data or Independent Analysis.....	50
2.	The Choice of a Specific Measure.....	52
a.	The Construct and Selection of Specific Measures.....	52
b.	Data and Variable Transformations.....	54
3.	Evaluation Criteria.....	56
a.	Ex Ante Criteria.....	56
b.	Ex Post Criteria.....	58

E.	MODELING TECHNIQUE.....	59
1.	The Issue.....	59
2.	The Literature.....	61
a.	Univariate Discriminant Analysis (UDA).....	61
b.	Multivariate Discriminant Analysis (MDA).....	61
c.	Conditional Probability Models.....	62
d.	Recursive Partitioning (RP).....	64
e.	Index Models.....	64
f.	Artificial Intelligence (AI).....	65
g.	Other Techniques.....	65
3.	What We Know.....	66
F.	VALIDATION.....	66
1.	Fit of the Model.....	66
2.	The Performance of the Model.....	69
a.	Performance On a New Sample.....	69
b.	The Costs of Errors.....	72
G.	CONCLUSIONS.....	73
1.	Theoretical Basis.....	74
2.	Sample Selection and Data Collection.....	74
3.	The Dependent Variable and Definition of Failure.....	74
4.	The Independent Variables.....	75
5.	The Modeling Techniques.....	75
6.	Validation.....	76
H.	RECOMMENDATIONS FOR FURTHER RESEARCH.....	76
V.	IMPROVING DEPARTMENT OF DEFENSE FINANCIAL ANALYSIS.....	79
A.	REVIEW OF CURRENT DOD PRACTICES.....	79
B.	DEFENSE INDUSTRY SPECIFIC RESEARCH.....	79
1.	DoD Specific Models.....	80
2.	Other Relevant Studies.....	83
3.	Research Evaluating the DoD Specific Models.....	83
C.	CONSTRUCTING A BETTER DOD FINANCIAL SCORING MODEL	85
1.	Theoretical Basis.....	85
2.	Sample Selection and Data Collection.....	86
3.	The Dependent Variable and Definition of Failure.....	87
4.	The Independent Variables.....	87
5.	Modeling Technique.....	88
6.	Validation.....	89
D.	CONCLUDING REMARKS.....	90
	LIST OF REFERENCES	93
	INITIAL DISTRIBUTION LIST.....	99

LIST OF FIGURES

1.	Sample Selection Considerations.....	10
2.	Independent Variable Considerations.....	14
3.	Z-Score Linear Classification.....	21
4.	Example of Recursive Partitioning.....	23
5.	Model Validation Considerations.....	25
6.	Cumulative Probability Density Function.....	64

LIST OF TABLES

1.	Bankruptcy Filings by Business Type.....	6
2.	Sample Size of Selected Failure Prediction Models.....	39
3.	Theoretical Basis for Independent Variable Consideration.....	57
4.	Modeling Techniques Used in Failure Prediction Models.....	60
5.	Classification Accuracy Rates (in Percentages).....	68
6.	Classification Accuracy of Retesting.....	71
7.	Dimensions of Financial Condition of DoD Contractors.....	83

ACKNOWLEDGEMENTS

The expert advice and guidance provided by my advisors, O. Douglas Moses and Shu S. Liao, were invaluable and deeply appreciated. This thesis would not be half its quality without their input. They taught me more than how to write a thesis, they taught me how to problem solve.

The remaining faculty in the Department of Systems Management deserve recognition for providing the critical tools necessary to approach this thesis, future Navy assignments, and future scholastic endeavors.

My children deserve special recognition for putting up with Daddy spending all of his time in the library and in front of the computer. They were not always understanding, and I thank them for teaching me how to write under duress. You can have the computer now, Kip.

My lovely bride, Ilene... your support was vital, your dedication was unwavering, and your love is cherished. Now I'm mighty smart, too.

And most of all, I thank God for the talents He has provided me, the richness of my life, and His overwhelming love.

I. INTRODUCTION

A. THE RELATIONSHIP BETWEEN THE DEPARTMENT OF DEFENSE AND THE PRIVATE SECTOR

The mid-1990s are a turbulent time for the relationship between the private sector and the federal government, particularly the Department of Defense (DoD). Budgets are declining in this post-Cold War era and in response to demands for a balanced budget and fiscal responsibility. Restructuring, Total Quality, Rightsizing, Base Realignment and Closure Commissions, Acquisition Reform, and the National Performance Review are dramatically affecting the way business is conducted. In this environment, DoD is continually seeking better ways to manage resources. Some initiatives in this vein include: contracting for value rather than simply price; strategic business partnerships such as the prime vendor programs for medical supplies and foodstuffs; and privatization of formerly, but not inherently, governmental functions such as base security and purchasing.

A January 1996 online search of the Reuters news service for keywords "privatization" and "outsourcing" yielded dozens of articles identifying new and expanded use of the private sector to perform government functions. Examples include:

- IVF America is under contract to perform assisted reproductive technology services in military hospitals and clinics;
- American Products Co. is providing food, trash, and fuel service for U.S. troops deployed to Hungary in support of Bosnian peacekeeping efforts;
- Bergen Brunswig Corp. is a medical-surgical supply prime vendor for DoD;
- Computer Science Corp. is providing acquisition support to the General Services Administration;
- C.I. Travel Company is providing travel reservation services for the U.S. Air Force;
- Rockwell is leading a team of seven other defense contractors in the private management of Air Force's Aerospace Guidance and Meteorology Center; and
- The proposal to privatize the U.S. Naval Petroleum Reserves.

The defense industrial base is concurrently undergoing rapid change in structure. Mergers and acquisitions are creating companies which may wield widespread influence on the shape of the military of tomorrow. Examples include: the mergers of Lockheed and Martin Marietta to create Lockheed Martin in 1994 and then Lockheed Martin's acquisition of Loral; Northrup merging with Grumman in 1994 and then purchasing Westinghouse's defense electronics division; and Raytheon acquiring E-Systems. Meanwhile other companies are restructuring to capitalize on private sector growth and are diminishing their dependence on the federal government. Two examples are AEL, Inc announcing that they are "actively pursuing commercial applications for its extensive and unique technologies" and GM Hughes Electronics shifting from military satellites to the commercial home system

called DirecTV.

Add to this turmoil, changes within DoD. First is the rapidly reforming acquisition process, highlighted by the Defense Acquisition Workforce Improvement Act, changes to the Cost/Schedule Control Systems Criteria system outlined in the recently revised DoD Instruction 5000.2, and the 1994 Federal Acquisition Streamlining Act. Second, is the decline in the Procurement, Research & Development, and Operations & Maintenance budgets. Third, is the shrinking size of the workforce available to administer the contracting function.

These trends in restructuring, strategic partnerships, revised business practices, and privatization are expected to continue. They have not only changed the financial condition of the businesses, but the government's dependence upon these businesses is increasing. To assure consistent long term support in a privatization contract or life cycle support in a weapon system procurement, the contracting agencies and program management teams must have the tools necessary to rationally evaluate the strength of firms competing for such work. Likewise, senior government officials must have reliable, accurate information regarding the state of the military-industrial infrastructure for policy decisions. This requires a methodology for determining the fiscal health of firms engaged in business with the government. One tool for assessing the fiscal health of a business is a financial scoring model.

B. FINANCIAL SCORING MODELS

A financial scoring model is a tool for financial analysis, designed to predict the likelihood of a firm failing based upon an analysis of data determined to have some statistical relationship with failure. It is best understood by examining how it is constructed. Normally two groups of data are compiled, one for a set of failed firms and one for a set of nonfailed, or healthy, firms. A set of variables is selected which are suspected of predicting failure, are descriptive of a firm's financial condition or (as is usually the case) some combination of both. Data is collected for those variables for each of the firms in the two sets. A statistical technique designed to discriminate between groups of data – in our case, failed firms versus non-failed firms – is selected and applied to the data, creating a model or equation. When data for a firm is entered into the model, the output of the model will indicate whether or not the firm is expected to fail or remain healthy.

An example is provided to illustrate. The most commonly cited model in the academic literature, and one still used within DoD, is a model developed by Edward I. Altman (Altman, 1968). It is in the form of an equation, the result of multidiscriminant analysis, and constructed of five financial ratios used as predictive variables. When data for a firm is entered into the model, a "Z-score" is computed for which an appropriate cutoff score is determined: a score above the cutoff indicates a healthy firm, a score below indicates a failing firm.

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \quad \text{Eq. 1}$$

where X_1 = working capital / total assets

X_2 = retained earnings / total assets

X_3 = earnings before interest and taxes / total assets

X_4 = market value of equity / book value of total debt

X_5 = sales / total assets

Although the use of mathematical tools to predict future events has been recorded as early as the 14th century, modern arithmancy, specifically the use of financial scoring models to predict business failure, began with the publication of Beaver's "Financial Ratios as Predictors of Failure" in 1966. Since then, numerous works have been published using increasingly sophisticated statistical techniques including univariate and multivariate discriminate analysis, probit and logit regression, recursive partitioning, indices, and artificial intelligence. Both numerical and non-numerical measures have been used as independent predictor variables in these analyses. Numerical variables have included traditional financial ratios, specialized financial ratios, trend analysis, and analysts' earnings estimates. Non-numerical variables have included the quality of wording in the annual reports, changes to accounting principles used in the preparation of financial statements, changes in management, and economic conditions.

With increasingly detailed databases from which to draw information and increasingly powerful tools for computation, the body of work is branching into many directions. Compounding the issue is that there does not exist a universally accepted theory of business failure; Chapters III and IV will cover this issue in more detail. A comprehensive look at the state of the art of failure prediction models has not been published since Zavgren (1983) and Jones (1987), yet significant research has been published since then. There appears to be a legitimate need for a new compilation of the body of work on the prediction of business failure and an evaluation of that work across a broader framework than that used by the most recent review (Jones, 1987). As noted in the next chapter, some DoD activities still use a financial scoring model developed in the late 1960s. Given the DoD's increased reliance on the private sector, there also is a need for an interpretation of the literature for a DoD perspective.

C. RESEARCH QUESTIONS AND LIMITATIONS

The primary research question to be answered by this study is: What is the current state of the art in the use of financial scoring models for the purpose of predicting business failure? More specifically, the study will review, summarize, evaluate, and critique the literature related to existing financial scoring models used to predict business failure.

The research for this thesis will consist of the identification of recently developed models; comparison of the models within a framework developed by the author and based upon the literature related to failure prediction; evaluation of the literature based upon the

comparison; and, finally recommendations for use of financial scoring models in various applications. Particular attention will be paid to the application of such models in evaluating the viability of defense contractors, those engaged in both procurement of systems and privatization of infrastructure.

The study is primarily a literature search, compilation, and critical evaluation of the existing body of work on the prediction of business failure. The study does not intend to test the researched models with new data, rather they will be compared and evaluated based upon the merits of the original research and any published criticisms of that research. That is, if a subsequent article has tested a model with new data, the original model and results will be evaluated and consideration will be given to the criticism as well. The study will compare and contrast the various statistical techniques employed, their underlying assumptions, and the strength of the models given those assumptions. However, given the complexity and variety of different techniques, this portion of the methodology must be succinct. The aim is to comment on their use in the context of evaluating the models.

D. ORGANIZATION OF STUDY

Chapter II of the thesis will consist of background material. It will begin with a look at the characteristics of firms that fail, answer the question of why failure should be studied, and the state of financial analysis within DoD. In Chapter III, a framework will be developed as the basis for comparison and critique of financial scoring models and an assessment of the literature related to failure prediction. This framework will expand on those used previously in the literature and will include (1) the theoretical basis for the model, (2) sample selection, (3) the dependent variable and definition of failure, (4) the selection criteria and quality of the model's independent variables, (5) the modeling technique used, and (6) validation approaches. Chapter IV will evaluate the literature within this framework with the goal of providing a snapshot of the state of the art. This chapter will also include concluding remarks and recommendations for further research in the broad academic context. Chapter V will narrow the focus to a DoD context and provide an examination of the subset of the literature which is specifically related to the use of models in a DoD context. The goal of this chapter is to evaluate the DoD literature against a backdrop of the broad academic literature and to make recommendations for improving financial analysis within DoD.

II. FOUNDATION FOR ANALYSIS

This chapter will summarize the literature regarding causes and conditions of business failure and the characteristics of firms that fail. Next, the use of financial scoring models within the Department of Defense will be discussed. Finally, some defense-specific research regarding financial scoring models will be introduced. This chapter will provide the foundation for the analysis to follow in subsequent chapters.

A. CHARACTERISTICS OF FIRMS THAT FAIL

To predict the event of failure, one must ask what failure is. Definitions previously applied in the private sector have included: negative working capital, court supervised reorganization and protection from creditors (bankruptcy), private asset and financial restructuring, bond interest default, preferred stock dividend default, and complete liquidation. In a DoD context, failure could be reflected in non-performance on a contract or financial distress severe enough to require prepayment of a contract, but not severe enough to result in bankruptcy. Filing for protection under Chapter 11 of the bankruptcy code is the definition most often cited in the literature.

With respect to bankruptcy, what is known is that firms file for reasons of insolvency, reorganization, or even to avoid labor disputes; they enter voluntarily or involuntarily. Claimholders who influence the business to file for bankruptcy protection include equity holders, bond holders, banks or other lending institutions, and trade creditors. These influences are normally asymmetrical and claimholders will behave in such a manner as to maximize their own outcome, if necessary at the expense of the others.

Dickerson and Kawaja (1967) studied the failure of firms from 1920-1965 on the basis of economic cycles, regions, industries, age of firms, and size of firms. They reached the following conclusions regarding the causes of failure and characteristics of failed firms:

- failure rates vary roughly in accordance with business cycles;
- failure rates vary among lines of business with the retailing sector showing the highest failure rate;
- firm life expectancy increased with age; that is, the longer a firm is in existence, the longer it is expected to remain in existence;
- firm size and failure rate are inversely related;
- managers with any prior experience running a business were more likely to have their businesses survive than first-time managers;
- the more capital invested and the higher the equity-to-debt ratio, the lower the failure rate;
- the failure rate is inversely proportional to the age of the manager;
- management teams were more successful than single managers.

We also know something about the rate at which firms fail. Table 1 clearly demonstrates that, recently, the firms most likely to fail are private, and most probably, smaller firms. Less than one-fifth of one percent are publicly traded firms and less than half of them have publicly traded bond debt.

	1990	1991	1992	1993	1994
Business Bankruptcy Filings	64,853	71,549	70,643	62,304	52,374
Public Company Bankruptcy Filings	115	125	91	86	70
Public Companies with Publicly Traded Bond Debt	37	58	42	32	26

Table 1. Bankruptcy Filings by Business Type

Within the literature are many studies which have demonstrated relationships between failure and some event or economic state, the actions taken by management when in financial distress to minimize the possibility of bankruptcy, the incentives of claimholders and their potential actions, and the capital structure of firms experiencing financial distress. The literature has proposed and tested a variety of theories regarding the nature of failure and the content of information sets which may be useful in predicting failure. These theories will be covered in depth in Chapter IV.

The research has provided some information regarding the dynamics of firms entering, and their subsequent behavior in, periods of financial distress, yet a complete and widely accepted theory of business failure has yet to emerge. Given this fact, it is not surprising to see that the study of failure prediction has extended in varied directions with little consistency in methodology and selection of predictive variables. In fact, Platt (1985) outlines three methods for predicting failure. The first is based upon thirteen common sense indicators, eight company-specific signs and five product signs. The second is based upon the use of financial ratios which he groups into six taxonomies and then provides a "best" ratio for each. His third method of prediction is the use of financial scoring models which, of course, will be covered at length in subsequent chapters.

B. WHY STUDY FIRM FAILURE

When a firm under contract to the government fails, or declines to some state of financial distress, there can be high economic costs to the government. A firm in financial trouble may need advance payments on a contract to ensure sufficient cash flow to complete the contractual obligations. A firm may be unable to meet cost, performance, or schedule requirements of the contract. A firm may "cut corners" inappropriately. A firm may

liquidate, declare bankruptcy, or be forced to sell off necessary assets in order to survive, impacting their ability to meet contractual obligations.

There are also potential costs to the government if firms within the defense industry fail, even if they are not currently under contract. There may be reduced competition within the industry, resisting the ability to compete contracts or providing a surviving firm undue influence. Note the recent "blessing" by DoD to the Lockheed-Martin acquisition of Loral's defense unit prior to the Federal Trade Commission's approval.

Being able to foresee these events could save the government significant costs; costs not just in dollars and cents, but costs associated with misspent time, wasted manpower, delayed fielding of a system, and economic instability.

How big a problem is it? Table 1 demonstrates that over 150 firms fail every day and that the firms most likely to fail are private, and most probably, smaller firms. In fiscal year 1994, of the \$112.0 billion in prime contracts let by DoD, 22.1%, or \$24.8 billion went to small businesses (and one fourth of those went to disadvantaged small businesses). Table 1 also shows that less than one-fifth of one percent are publicly traded firms and less than half of them have publicly traded bond debt; of course, they receive the remaining 77.9% of contracts. Exact figures on the scope of the problem within DoD are unknown, but the circumstantial evidence indicates that the costs could be very high. Zmijewski (1984) stated that the failure rate in his population ranged from 0.49% to 0.94% each year and that this was consistent with failure data published by Dun and Bradstreet. Taking a mid range figure of 0.75% per year and multiplying by the \$112.0 billion in prime contracts indicates that \$840 million in prime contracts may be at risk each year.

C. FINANCIAL ANALYSIS WITHIN THE DEPARTMENT OF DEFENSE

Borah (1995) identified five activities within the DoD which are involved in financial analysis. Two of the activities, the Defense Contract Management Command (DCMC) and the Defense Contract Audit Agency (DCAA) conduct analyses in support of the contract award process. The remaining three activities, the Naval Center for Cost Analysis (NCCA), the Army Center for Resource Analysis and Business Practices (ACRABP), and the Air Force Office of Economic and Business Management (OEBM) conduct their analyses to support the Independent Cost Estimate and Cost and Operational Effectiveness Audit inputs to acquisition milestone reviews. These three activities are also tasked by the service secretaries to assess the financial health of their respective service's industrial base.

In performing their analyses, each command takes a unique approach, some relying on financial scoring models, others taking a more qualitative approach. DCMC has published the Guide to Analysis of Financial Capabilities for Pre-award and Post-award Contracts which prescribes a standard form to be completed by the analyst. The form provides for a summary of the firm's balance sheet and income statement and the analyst

computes three financial ratios, all measures of liquidity. The guide recommends, but does not require, the use of Altman's original model (Altman, 1968).

The DCAA Contract Audit Manual (§4-804.4) discusses the indicators of solvency problems and provides for the use of Altman's model.

Failure prediction models in general provide a means to readily assess a contractor's financial health in terms of the likelihood of bankruptcy in the near future. Therefore, the auditor should analyze the contractor's financial data by means of a financial failure prediction model.

The model results are augmented by an examination of bank lines of credit, liquidity ratio analysis, cash flow forecasts, operating profits or losses, and various other items of a qualitative nature (e.g., labor disputes, unusual audit opinions, and contractor plans for dealing with adverse business conditions).

The NCCA uses the model developed by Dagel and Pepper (1990) in conjunction with ratio analysis of liquidity, solvency, and profitability. ACRABP uses a combination of Moody's or Standard & Poor's bond rating services and ratio analysis. They use six ratios measuring solvency, four measuring efficiency, and three measuring profitability. OEBM has moved away from using financial scoring models and relies almost exclusively on Moody's and Standard and Poor's bond ratings. They place little emphasis on ratio analysis as it is inherent in the the bond ratings.

The techniques used by the various DoD activities vary by the intended use for the analysis and the service performing the analysis. While it seems logical that different applications would require different techniques, those commands applying their analyses in similar fashion use remarkably different techniques. One would expect more consistency. As Borah (1995) recommends, follow-on research should be conducted to determine which of these various techniques is most effective. The author hopes that this thesis will be helpful in improving the accuracy of failure prediction and leads to more effective use within DoD.

D. SUMMARY

This chapter has introduced the literature regarding causes and conditions of business failure and the characteristics of firms that fail. The need for studying the failure of firms and the costs associated with failure were introduced. Finally, the methods of financial analysis conducted within DoD were examined and, specifically, the use of financial scoring models. Of particular note is the diversity of techniques used to conduct financial analysis and failure prediction. Given this foundation, the next chapter will develop a framework for analyzing financial scoring models based upon the literature related to failure prediction and, for each dimension within that framework, issues to be considered when designing or applying the models.

III. FRAMEWORK FOR ANALYSIS

To assess the financial health of a business, one can take any of several approaches: commercial services (e.g., Moody's and Standard & Poor's), financial ratio analysis, economic forecasting, a qualitative examination of the business's relative market position and strategy, a financial scoring model, or some combination. As noted last chapter, Platt (1985) outlined three methods including "common sense" indicators. Of course, the best approach can only be determined by the user given the context of the assessment. But this user would need to develop some method of analyzing the relative merits of the various approaches. As the focus of this thesis is the financial scoring model, a framework for analysis of these models is developed here.

Jones (1987) evaluated the state of the art based upon a framework (inspired by Scott, 1981) consisting of four dimensions: sample selection, choice of independent variables, choice of statistical methods, and the evaluation of empirical results. Finding Jones' work incomplete, particularly in light of the expansion of the literature since his writing, the author will present in this chapter a framework consisting of six dimensions for analysis of financial scoring models and introduce the issues surrounding each. The issues are often complex, overlapping, and comprise choices and tradeoffs for the developer and user; each of the choices and tradeoffs has implications for model usefulness and validity. The six dimensions are: (1) the theoretical basis for the model, (2) the sample and data collection, (3) the dependent variable and definition of failure, (4) the independent predictor variables, (5) the modeling techniques employed, and (6) the validation process.

Subsequent chapters will use the same framework to evaluate models from the literature and the use of financial scoring models in a DoD context. While other approaches for evaluating financial scoring models have been used in the literature – chronological evolution, within a particular industry context, along a particular statistical technique – none has systematically evaluated the state of the art along all of the dimensions outlined below.

A. THEORETICAL BASIS

The scientific method suggests that hypothesis precedes experimentation which precedes analysis which precedes theory. Given an accepted theory, researchers may use its principles to conduct further experimentation to expand the body of knowledge. As discussed last chapter, a universally accepted theory of business failure has yet to emerge, confounding the development of models to predict such an event.

This is not to say that this area of research is completely devoid of theory. Two classes of theories are potentially relevant to financial scoring models. The first class are models whose content is based upon theories of firm behavior. The second class are models that rely upon a theoretical basis for the selection of predictor variables, i.e., theories regarding the content of various information sources, such as financial ratios.

B. SAMPLE SELECTION AND DATA COLLECTION

The selection of a sample and source of data from which to draw that sample are particularly problematic when constructing a model for the purpose of predicting failure. Issues fall roughly into two categories, conceptual issues surrounding the content of the sample, and practical issues surrounding the composition of the sample and collection of the data. Figure 1 shows these issues graphically and they will be discussed in turn below.

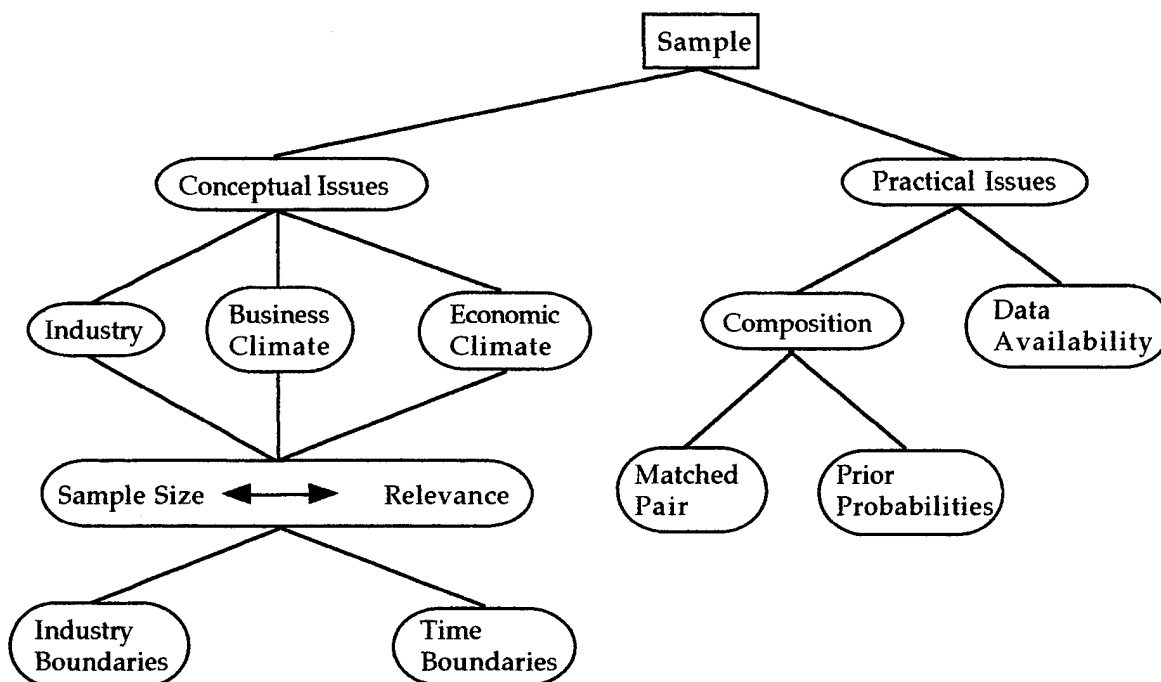


Figure 1. Sample Selection Considerations

1. Conceptual Issues

The conceptual issues raised and tradeoffs encountered relate to the questions surrounding the context of the model: the business environment, industry, and economic conditions; the inherent tension between a sample's size and its relevance, and finally the tradeoffs with respect to industry segments and time periods sampled.

a. Industry, Business Climate, and Economic Climate

Conceptually, the developer or user of the model must first address three issues: the industry or industries being studied, the business climate, and the economic climate. The industry under consideration is the first of these. At stake is how broadly applicable the model will be to various businesses. Boundaries may be placed on business size, output, or customer; a useful gauge is SIC code from the *Moody's Industrial Manual*.

Another issue related to industry is the uniformity of the "message" predictor variables provide. For example, a healthy financial services business will have very different values for its financial ratios than an equally healthy industrial or retail business.

The business climate, second of the three issues, is a particularly important consideration when developing a model for DoD. Recently, the post-Cold War era has reshaped the defense industry. Reduced procurement and operating budgets in the 1990s, following the rapid expansion of the 1980s, have shocked the industry. Ramifications have included privatization of governmental activities, mergers and acquisitions within the defense industry, and a shift from native research and development to the use of more commercial off-the-shelf components.

The third conceptual issue is the economic climate. The question at issue is whether the model can be validly applied during economic expansionary, recessionary, and stagnant periods. The researcher must decide if the model intends to be robust across all climates or whether it is specifically designed for application in a particular climate, (e.g., to predict failure during economic downturns).

b. The Tension Between Sample Size and Relevance

Consideration of the three conceptual issues above raises a larger issue: the developer faces a trade off between sample size and the relevance of the model. The rate of failure among businesses in the United States is extremely low, yet a sufficiently large sample size is needed to formulate and validate a useful model. Drawing a large sample size can risk compromising a specific relevance of the model since the boundaries of time or industry must be expanded. On the other hand, drawing a small, more specific sample will risk compromising the statistical relevance of the model. When drawing the sample, the two dimensions at issue are the industry boundaries and time boundaries. They will be discussed in turn.

c. Industry and Time Boundaries

As noted above, of particular concern in the selection of a sample is the infrequency of failure. To have sufficient data for a statistically significant model, the sample must either encompass a diverse number of industry segments or a long period of time. Either choice has its drawbacks. Use of data across varied industries is likely to confound any industry-specific patterns among the predictor variables; this is particularly problematic when combining service sector firms with industrial firms and when including firms in the financial sector. The use of intra-industry relative ratios can alleviate some of these problems (Lev and Sunder, 1979). However, the researcher's context may preclude crossing segments, as may be the case in developing models for assessing the health of a particular industry. A sample may then need to be drawn from a lengthy period of time.

Time introduces other problems. The macroeconomic climate and business cycles may change, there may be changes to accounting principles affecting comparability, and emerging technologies may have affected the business climate. All of these problems

may lead to a suboptimum model. On the other hand, a lengthy period of time would be useful should the researcher be using a time series dependent predictive variable.

2. Practical Issues

In addition to the conceptual issues surrounding sample selection, there exists two practical issues: the composition of the sample across categories and the availability of data.

a. Composition

The researcher must decide on the composition of the sample. The two principal choices are whether to construct a sample based upon matched pairs of failed and non-failed businesses or to approximate the relative proportions which exist in the population. The relative proportions, while truer to the population demographics, normally necessitate an enormous sample size in order to obtain enough failed firms, and the proportion of failed firms grows smaller as the firm size increases and ownership is public. Matched-pairs, however, require careful matching on the part of the researcher to ensure the two halves are indistinguishable except for the failure event. This matching removes the randomness from the selection process and adds a bias to the results. Depending upon the criteria for matching (normally size of the firm and industry), other biases or distortions can be introduced. Joy and Tollefson (1975) and Zmijewski (1984) discuss these biases at length.

b. Data Availability

The source of the data may also introduce bias. Data is available from commercial sources such as the *Compustat* file, *Moody's Industrial Manual*, and the *Wall Street Journal Index*; governmental sources such as the Securities and Exchange Commission; or the business itself. One must question the filtering mechanisms imposed by the data source, i.e., the segments of the population which are included and excluded by the data source and what biases this introduces. Jones (1987) specifically mentioned the problem of data for smaller firms, noting that only those which were deemed "newsworthy" would have sufficient data available. If multiple sources are used, the data should be comparable and consistent. These issues are especially problematic when the research involves privately held firms. Many of the commercial and governmental sources will have sparse, if any, data.

For a thorough look at the changes in financial health of a firm which eventually fails, it is necessary to look at several years worth of data. Commercial sources may not have complete databases for all firms under consideration, particularly for those which have recently gone public, further limiting the sample size. Young firms may also be excluded due to a lack of sufficient data, despite the fact that they are more likely to fail.

Another dimension of sample selection involves the intended validation technique. If the researcher intends to validate the results on a sufficiently large hold-out sample, often that necessitates aggravation of the problems cited above regarding crossing

industries and lengthy time periods; see Section F of this chapter for more information about validation samples.

C. DEPENDENT VARIABLE

Similar to the case of the sample, the dependent variable raises both a conceptual and an operational issue for the model developer. The conceptual issue relates to the question of the construct under study. The operational issues relate to the question of the scale used to measure the output. As stated in the first section of this chapter, some models are based upon a theory of firm behavior or movement into a particular state of financial health. These theories will affect the choice and composition of a dependent variable.

1. The Construct Being Investigated

The choice of a dependent variable goes to the construct being investigated, the definition of failure in use. This has ramifications throughout the model's construction: the sample chosen must come from a sufficient and relevant population; the variables selected as predictors will be affected; and even the best statistical technique to use is a function of the definition of failure.

Of particular note is the use of bankruptcy as the dependent variable: Dietrich (1984) points out that "bankruptcy is a legal, rather than an economic, condition." He asserts that the use of an economic model will be limited in its ability to predict a legal event. Jones (1987) reinforced the point. Others may choose to construct a model, not with failure as the dependent variable, but rather health. An analysis of a firm's potential to fail may just as well be determined using a model designed to indicate health as well as one designed to indicate failure.

2. The Scale of the Output

The researcher must decide whether to use a discrete or continuous measure. At one end of the continuum is a model designed to predict a unique event, e.g., bankruptcy, which would thus use a dichotomous dependent variable. The model would provide an output that indicates failure or nonfailure. A model can also be designed such that the dependent variable provides for more than two discrete states. Such a polytomous model can be very useful in some applications: perhaps a user wishes to assess several degrees of failure or states of financial distress. On the other end of the continuum is a variable which provides for a continuous distribution of outputs. These outputs can be either discrete values along a continuous scale for which cutoff scores must be assigned for classifying firms, or a probability estimate of failure.

D. INDEPENDENT VARIABLES

If there existed proven theories of failure, the selection of independent variables would be relatively simple. Given the lack of an accepted theory to guide the selection of

variables, this is the one dimension over which the most intuition, logic, statistics, and creativity can be exercised. It is also the dimension which arguably has the most impact on the quality of the subsequent model. As previously discussed, the knowledge about business failure is sketchy and each failure is unique in its causes and evolution. The developer of a model has many choices to make and issues to confront. They will be discussed systematically, first, along the broad issues of the information set; second, the selection of specific measures to represent the information set chosen; and third, criteria to be met in evaluating the quality of the specific measures. These issues are presented graphically in Figure 2.

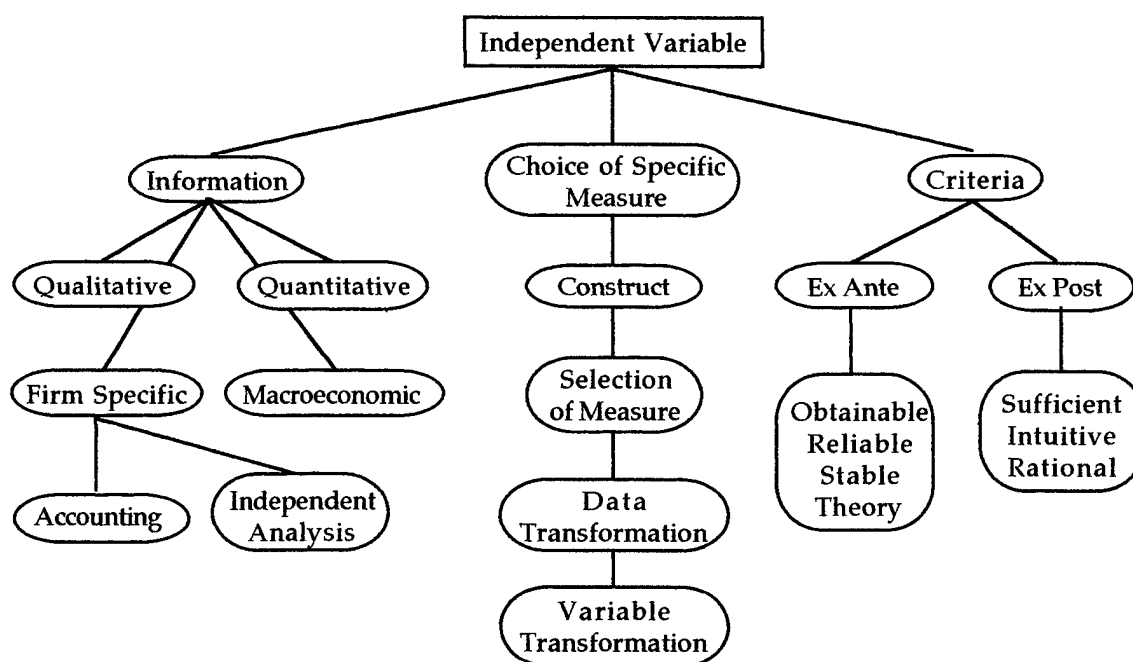


Figure 2. Independent Variable Considerations

1. The Information Set

The information set is what normally comes to mind when one envisions a financial scoring model. What are the indicators of financial health or failure? Ideally, the decision of what information set to use should be driven by theory. Again, failure theory has been sparse and some researchers have instead used existing theory of the behavior of certain predictive variables (most commonly, financial ratios) to determine the information set. Other researchers have simply used statistical techniques to select specific measures from a large data set without particular regard paid to the information content of those choices. A

well-conceived model should address the information content first, then choose appropriate measures of that information.¹ There are a few fundamental choices the model developer faces when building the information set.

a. Qualitative and Quantitative Variables

First, the variables will assume a qualitative or a quantitative aspect. (In this discussion, a qualitative variable is synonymous with a dummy variable: information which is not readily converted to a numerical value and is incorporated in the model by using a (0,1) convention, 0 representing the absence of the indicator, 1 its presence.) Qualitative variables will be addressed first. They generally convey information about a business beyond what the numbers reveal, tending to be forward-looking rather than records of historical achievements. They also can better convey such constructs as the organizational complexity of the business.

Senge (1990) discusses two types of organizational complexity: detail complexity and dynamic complexity. Detail complexity is the act of distilling an organization into component parts to determine cause-and-effect. Dynamic complexity recognizes that organizational states result from intricate interactions among various processes working within and without the organization. An argument can be made that failure is a dynamic process and would suggest that the use of qualitative variables would be particularly useful. Giroux and Wiggins (1984) develop an events approach framework for failure in general and bankruptcy in particular. They found that "the events most closely associated with bankruptcy are net losses, debt accommodations, and loan default," the latter two comprising qualitative variables. Hawkins (1986) states that 30-50% of a bond rating is attributable to "management, industry, general economic conditions, future prospects, and other qualitative factors." Other events which may be milestones in the failure process include: changes to accounting principles which give the illusion of higher income or improved cash flow, incidences of management turnover, asset sales, downgrading of bonds, financial restructuring, and the reduction or elimination of a common stock dividend. Such ideas derive from the works of Schwartz (1982), Matthews (1983), Wruck (1990), DeAngelo and DeAngelo (1990), John, Lang, and Netter (1992), Ofek (1993), Opler and Titman (1994), and Khana and Poulson (1995) and will be discussed in detail in Chapter IV.

On the other hand, the variables could be of a quantitative nature. The most common quantitative measure is a financial ratio. Traditional financial ratios have strong appeal; they are intuitive, reliable, and easily obtained. There is an extensive literature to draw from regarding the behavior of financial ratios. Research has been done on the stability of ratios across industries and economic cycles, the tendency of individual firm's ratios to

¹ Of course, a pure statistical reduction may be the researcher's intent. Some research has been done using factor analysis, stepwise inclusion and reduction, and other techniques with the goal of simply providing information about the descriptive nature of the data. The point the author is making is that research of this type should be presented as discovery and not a model to be applied in practice.

move toward industry means, and the development of taxonomies of financial ratios. Works in these areas include Lev (1969), Pinches, Mingo, and Caruthers (1973), Libby (1975), Dambolena and Khoury (1980), Chen and Shimerda (1981), and a recent DoD specific work by Moses (1995). Additionally, non-traditional financial ratios may provide some insight into the dynamics of firm failure.

Lev and Sunder (1979) raise an interesting issue with respect to financial ratios. Ratios are commonly used to control for size or some other industry wide factor, yet size has been shown to be highly correlated with failure. The researcher must ask if it is prudent, therefore, to deflate its effect, and if so, whether a ratio is the proper technique for doing so.

b. Firm Specific or Macroeconomic Variables

The next question is whether the variables, qualitative or quantitative, provide information about the firm specifically, or some macroeconomic set, i.e., the industry or entire economy. It may seem logical that a firm specific indicator is both necessary and sufficient. In fact, as we will see, most developers of models hold this opinion. The rationale is that in a well-selected sample, the broader economic effects are felt by all businesses and they do not uniquely affect the failure event for a particular one. The information, the argument continues, would be incorporated in the firm's specific data. For example, in a recessionary period marked with high inflation, one would naturally expect the ratio of cost of goods sold to total sales to rise. Higher prices would lead to higher costs of production and reduced demand for the firm's product.

Depending upon the context of the model, macroeconomic indicators, however, have an intuitive appeal. The rate of business failure tends to correlate with both broad and industry-specific economic cycles (Rose, Andrews, and Giroux, 1982). If the overall probability of failure is greater during poor economic conditions, then the probability of an individual firm's failure would rise correspondingly. This is logical: a high cost of capital during inflationary times would make a cash-flow poor distressed firm more likely to fail, likewise a heavily indebted business would be at a disadvantage expanding capacity during a boom period, potentially leading to its failure.

c. Accounting Data or Independent Analysis

When using firm-specific data, the developer has one additional decision regarding the information set: whether the model will employ the firm's accounting data or rely upon independent information regarding the firm. Certainly accounting information is readily available, particularly for large, public firms. But it can be problematic for small, privately-held companies. Accounting data is also pure since it has not been modified, supplemented, or filtered by the information source and has been independently audited. Given proper disclosure and adherence to generally accepted accounting principles, the data is reliable and comparable to other businesses.

There are also benefits to independent analyses, e.g., capital market information, bond rating services, and security evaluation services. Capital market

information may be beneficial in that some level of analysis of the business is inherent in the data; analysis which recognizes interactions between qualitative data and financial measures. Given a reasonably efficient market, stock prices represent the perceived future prospects of the business. This is calculated within the market, requiring no additional work on the part of the user. The problem is one of knowing what the market really expects: the stock price could be reflecting the expected net present value of the future shareholder returns of the business, or the expected liquidated value in the event of failure, or the acquisition value, or some probability distribution of all three. The question is whether this uncertainty can be reliably used in the model. Commercial bond rating services perhaps provide better insight into the financial health of the business. With a multi-stage scoring system, the user of the data has some indication of the relative strength of the business. Table 1, however, showed a serious deficiency in using capital market information: the failure rates of businesses which are public (and, particularly those that issue publicly traded debt) are very small requiring a sample to be taken over a relatively long period of time. Another problem is that that data are not obtainable for smaller firms or firms with a different capital structure.

2. The Choice of A Specific Measure

Once the information set is determined, the next issue is the selection of specific measures within that information set. There is a hierarchy which should be followed when selecting measures; the reader is referred back to Figure 2. Each step in the hierarchy narrows the search for good variables, ensuring statistical significance. It is also necessary to follow the hierarchy to ensure the selection of variables which meet the evaluation criteria discussed in Section 3 below.

a. The Construct

Having selected an information set, the next consideration is the construct. In other words, the construct is dissected in such a way that the independent variables compose the most representative set to describe that construct. This may be best explained by use of an example. Suppose a researcher intends to predict business failure using a theory of cash flow and the information set chosen is a quantitative one using firm specific financial ratios. The construct issue relates to the questions: how does one measure cash flow? and what comprises cash flow? The answers lead the developer to the selection of specific measures.

b. Selection of a Measure

Continuing the example, the researcher must determine proper measures for cash flow. Dissecting cash flow, one of the issues is that of liquidity. The researcher then searches for an appropriate measure of liquidity. A common one is the current ratio, the firm's current assets divided by current liabilities. This selection is made based upon prior research, the underlying theory of the variable set, and the previous step, the construct of the variable. At this point, it may be desirable to select more than one representative measure and test each, discarding those that are not statistically significant.

The developer must be careful to ensure the reduction technique does not overfit the data and become less effective in application. Overfitting occurs when the statistical reduction technique is performed rigorously on the sample data such that it describes the sample precisely, but loses effectiveness in general application with other data. Eisenbeis (1977) cautions "that it may be unwise to drop dimensions or variables without first exploring in more detail what the possible effects may be." And Altman, et. al. (1981) concluded that "variable reduction techniques are not to be used as a substitute for theory or to derive underlying causal relationships or models."

c. Data Transformations

Once the specific measures have been selected, the developer must look at the information source to determine if there are any required data transformations. Transformations may be necessary for several reasons. First, is the notion of comparability. All data must be consistent and derived in the same fashion in order to be comparable. Incomparability may result from different accounting principles in practice (e.g., inventory valuation or asset amortization), the age of the data with respect to the failure event, or biases introduced by differing sources of information.

The age of the data can be particularly problematic when constructing a failure prediction model. The developer must be cognizant of the date of the public release of the data with respect to the date of the failure. Certainly, data released after the failure event is of no predictive value. But what of data released a few weeks prior to the event; can it be legitimately compared to data released a few months prior to the event? Ohlson (1980) noted that businesses facing failure took longer to release financial statements than healthy firms, often delaying the release of statements until after the failure announcement. The earliest statements available prior to the failure event averaged thirteen months in age.

The second basis for transformation results from the need for relevant data. If macroeconomic conditions are at issue, for example, the data may need to be transformed to reflect the effects of inflation or credit policies. If the sample crosses industries or industry segments, it may be necessary to transform the individual business statistics into an intra-industry ratio by dividing by the industry mean (Platt and Platt, 1991). This transformation would allow for the fact that a specific value for a variable may indicate health in one industry but not in another. Historical cost accounting for the value of fixed assets may also distort financial ratios when used to compare businesses, particularly if the businesses may be selling assets or spinning off subsidiary business segments. Jones (1987) raised a concern, however, regarding data transformations arguing that the usefulness of this technique may be lost if the distortion affects the meaning of the values.

d. Variable Transformations

The final decision facing the model developer in the selection of specific measures is whether any of the specific variables require transformation. A simple plot of the predictor variable against the dependent variable may indicate the need to mathematically transform the variable to create a linear relationship (if the modeling technique so requires);

such mathematical transformations include logarithms, square roots, and reciprocals. The modeling technique may require transformation, such as the multivariate normality assumption inherent in multivariate discriminate analysis. A transformation over time may be necessary if the variable is related to the failure event only if lagged by a certain period of time. Again, the developer of the model is cautioned that such transformations may distort the "message" sent by the variable.

3. Evaluation Criteria

Once the variables are selected, their quality should be assessed. Assessment should occur both prior to the model's development and then again after development. The first, *ex ante*, criteria are aimed principally at the content of the variable set; the second, *ex post*, criteria are aimed at their use within the model.

a. Ex Ante Criteria

Prior to actually constructing the model, the developer must look at the chosen variable set and assess it on the basis of whether it is obtainable, reliable, stable across the sample, and in conformance with the stated theory. These qualities will be examined in order. Obtainable refers to the ease and cost of acquiring the data in a useful form. A predictor should be readily available to the user of the model at nominal cost and it should not require excessive manipulation. Reliable data refers to the quality of the source. Audited financial statements, respectable commercial data sources, and governmental databases immediately come to mind. Qualitative variables are particularly vulnerable to unreliability, so a consistent method of evaluation must be used. The stability of the variables is especially problematic as the sample size grows beyond the boundaries of industry and time as different accounting rules, policy choices, and industry conventions are exaggerated.

b. Ex Post Criteria

Once the model has been developed, a second look at the quality of the variables is in order since during the construction of the model new issues will surface. The quality criteria, *ex post*, are sufficiency, intuitiveness, and rationality. A sufficient variable set must describe the essence of the failure event. There should be no statistically significant information missing, nor should there be excessive information present. Several statistical techniques are used to assure this quality such as stepwise inclusion and reduction, factor analysis, and F-statistics. The variable set should also be intuitive and rational. Intuitiveness questions the relative importance and the positive or negative sign assigned to the coefficient of the variable. For instance, if the model shows debt to equity as inversely related to financial distress, this may be statistically valid given the model's construction, but it is counterintuitive. Irrational variables can occur if the initial set of variables is large and the reduction technique is not monitored. With thousands of possible combinations of financial information, ratios may be developed which, for unknown reasons, have a high correlation with failure, but make no rational sense. This is not a farfetched notion. Researchers have

shown statistically significant correlations between the hemlines of skirts and stock market movement and between the calendar year and Presidential assassinations. While interesting, and even statistically valid, these relationships are either coincidences or aberrations of chance – the tails of the probability distributions – and should not be used to show causality or be used as a predictor.

4. Summary of Independent Variable Issues

There are three key issues to consider when evaluating or constructing the variable set: the information content, the selection of specific measures, and the quality of those measures. Special attention must be paid to the theoretical basis for the model, if any, and the construct under investigation. Prediction of the failure event depends upon a well-constructed variable set which contains relevant information and meets the ex ante and ex post criteria for quality.

E. MODELING TECHNIQUE

The selection of a modeling technique is the last step of the developer before validating the model. There exist several techniques designed specifically for discriminating data into two classifications. Each technique has its own assumptions, limitations, strengths and weaknesses. The techniques and their principle issues will be introduced here and expanded upon in the next chapter. The author recognizes Eisenbeis (1977), Altman, et. al. (1981), Collins and Green (1982), Zavgren (1985), Frydman, Altman, and Kao (1985), and Jones (1987) for their well-authored descriptions of the modeling techniques, synopsized below.

1. Univariate Discriminant Analysis

A univariate discriminant analysis (UDA), as the name implies, uses a single independent variable to classify a business into a failed or non-failed category. The technique normally involves the selection of several variables to consider individually and the computation of the data for the two groups. The next step is a comparison of the mean values of each variable for the two groups and the assignment of an appropriate cutoff score which discriminates between the two groups in such a way as to minimize the errors. The predictive ability of each variable is then applied to a holdout sample.

2. Multivariate Discriminant Analysis

Multivariate discriminant analysis (hereafter referred to as MDA) is a technique which categorizes businesses by a vector of multiple independent variables in such a way as to maximize the difference between the means of the categories when these multidimensional characteristics are mapped onto a one-dimensional measure (in our case, failure vs. non-failure). The technique generates a formula which provides the user with a “z-score” for

which a critical value is determined, a score above which indicates categorization of one type, below which, categorization of the other type. Figure 3, below, which is applicable to both univariate and multivariate models, illustrates the technique: the model generates a z score that maximized the separation between the group means, μ_i . The critical cut-off would be determined within the grey space, depending upon the cost of errors.

Although debated somewhat in the literature, one can use the z-score obtained by MDA to approximate the probability of failure if one assumes the z-scores are distributed normal. Criticism (Ohlson, 1980) of this manipulation charges that the normality assumption is invalid if the multivariate normality and equal covariance assumptions are violated. If one were interested in a probabilistic outcome from a model, the use of a conditional probability model seems more appropriate.

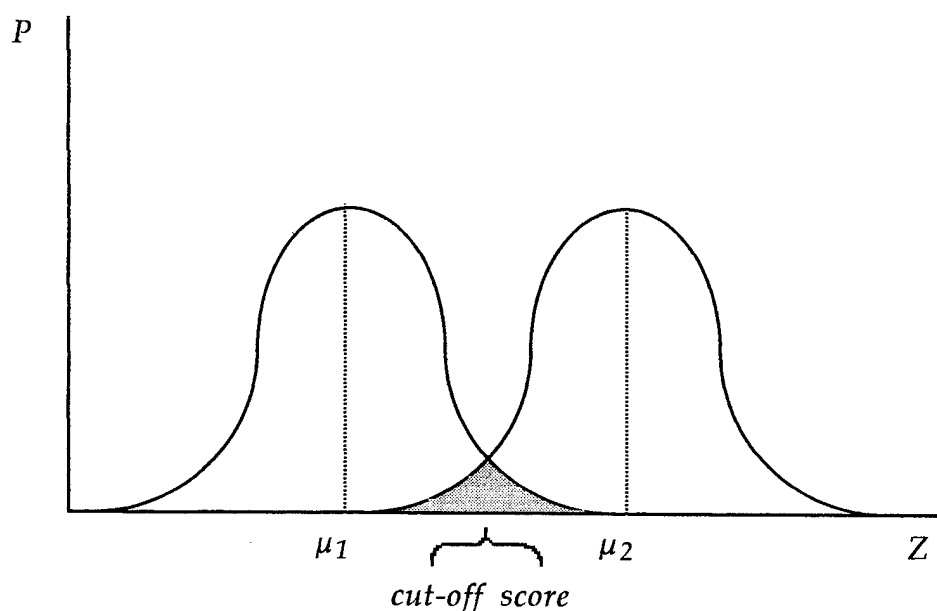


Figure 3. Z-Score Linear Classification

3. Conditional Probability Models

To provide a probability estimate of failure, and to avoid the restrictions inherent in the assumptions of MDA, a conditional probability model may be used. In essence, these models provide the conditional probability of a business belonging to a certain category (failed or non-failed), given the values of the vector of independent variables. As Ohlson (1980) described it in the first application of the technique to failure prediction, "The fundamental estimation problem can be reduced simply to the following statement: given that a firm belongs to some prespecified population, what is the probability that the firm fails

within some prespecified time period?" The underlying assumption is that the probability of failure can be described by the following formula:

$$P(\text{failure}) = \frac{1}{1 + e^{-X\beta}} \quad \text{Eq. 2}$$

where X is the vector of independent variables and β is the vector of coefficients weighted so as to maximize the joint probability of failure for failed firms and non-failure for healthy ones.

Two common types of conditional probability models exist, the probit and logit. Probit assumes the cumulative probability density function (cdf) is distributed normal, logit assumes a logistic cdf. Both cdf's have a mean, mode, and median at zero, but the logistic is more disperse: the standard deviation is 1.81 versus 1.0 for the normal cdf.

4. Recursive Partitioning

Recursive partitioning (RP) is another technique available for categorizing businesses as failed or non-failed. The principle benefit of the procedure is that it does not assume distributions for the independent or dependent variables, as do the MDA and conditional probability models. Another key benefit is its ability to minimize misclassification costs when the prior probabilities and costs of errors are specified.

Recursive partitioning requires the input of an original sample of data with their actual group categorizations as well as the costs of errors and prior probabilities. The model takes the form of an iterative binary classification tree: in stepwise fashion, the single independent variable which best discriminates the group into its categories at the lowest classification cost is selected and a cut-off score determined for assignment into each category. One of these two nodes is characterized by a preponderance of failed firms, the other would be more mixed. Each of these two nodes is similarly examined to determine the single best discriminating variable which minimizes the classification cost; two new nodes are established. This procedure continues until no further economic divisions are made and the tree ends with several terminal nodes. It is possible for a variable to reenter the "tree" as the divisions continue. Figure 4 provides a graphical example of the process. In the figure, F represents the number of failed firms, NF is the number of non-failed firms, X_i are independent variables, v_i are cut-off scores for those variables, and the dark circles represent terminal nodes.

The output is a classification matrix for which the costs of misclassification can be easily calculated. It provides some measure of the discriminating importance of the variables, however its forward selection process does not permit conclusions to be drawn about the relative importance of the variables or interactions between them. It also does not provide a probability estimate of failure.

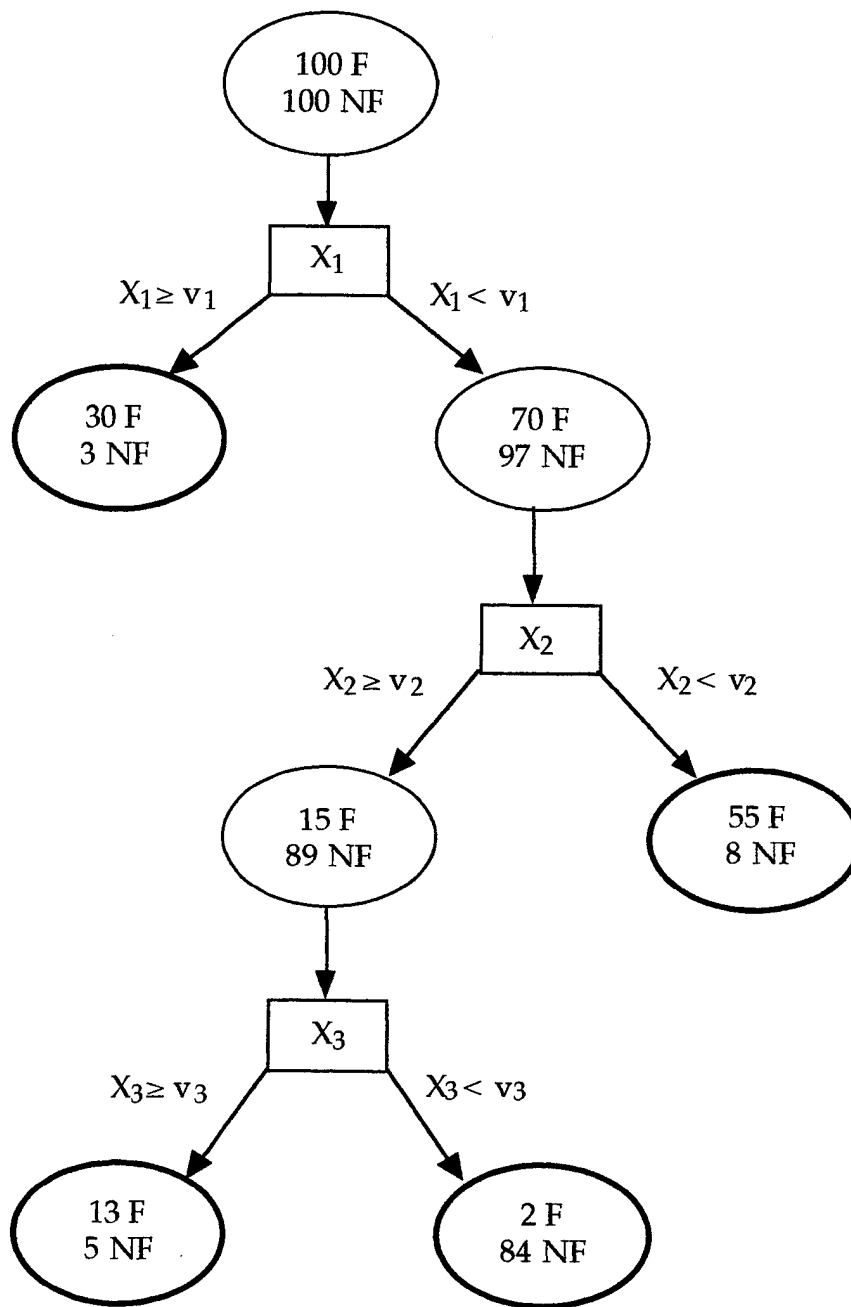


Figure 4. Example of Recursive Partitioning

5. Indexing

Indexing takes advantage of the simplicity of a univariate discriminant approach, yet addresses the issue of the multidimensional complexity of the financial status of a business. An index is derived by selecting N independent variables (ideally, based upon some taxonomy derived such that they capture all aspects of the financial condition of the business without introducing multicollinearity) and performing univariate discriminant analysis with each to obtain an optimum cut-off score. When this has been completed for all N variables, an index is constructed against which the businesses are scored. Scoring a business is done by comparing its data with the cutoff scores and assigning a score of 1 for each variable indicating failure, a 0 for healthy signals. An optimum sum total score is determined based upon the costs of misclassification errors.

6. Artificial Intelligence

As computational power has become dramatically less expensive and the knowledge base expands, more sophisticated methods of forecasting business events have evolved. In the field of artificial intelligence (also called "expert systems"), applications have included credit scoring, portfolio management, financial ratio analysis, personal financial planning, and tax advising (Coats, 1988). Weintraub (1989) suggests fourteen applications in government administration including "Forecasting – financial planning and cash management" and "Bid and proposal preparation assistance." The natural extension to business failure prediction has already begun.

What distinguishes artificial intelligence systems from the six modeling techniques outlined above is its ability, albeit limited, to mimic human reasoning and to "learn." In the most simple sense, these systems work by finding patterns in the reasoning and actions of a human analyst for a given set of development data, then use similar reasoning to score a new set of data. The computer will use the same data as the human analyst in conjunction with the analyst's decisions. Looking backward, the computer will decompose the logic process the analyst used to reach the conclusions. When new data are input to the program, similar reasoning is used, in a forward looking fashion, to determine the outcome. As Coats and Fant (1993) describe, the system "formalize[s] this ingrained, unarticulated knowledge of the experts by uncovering consistencies between the experts' conclusions and the recurring patterns in the financial data."

F. VALIDATION

The need to validate a model once constructed is not questioned. But there are several issues surrounding the best way in which to perform the validation. This section will address the validation issues using the framework in Figure 5, below. In short, there are two types of validations to be performed. First is a test of the statistical significance of the fit of the model. The second type of validation is the model's performance: how well does it discriminate between failed and non-failed firms? Performance is then be assessed in two

ways, through application of the model on different samples and through an analysis of the errors.

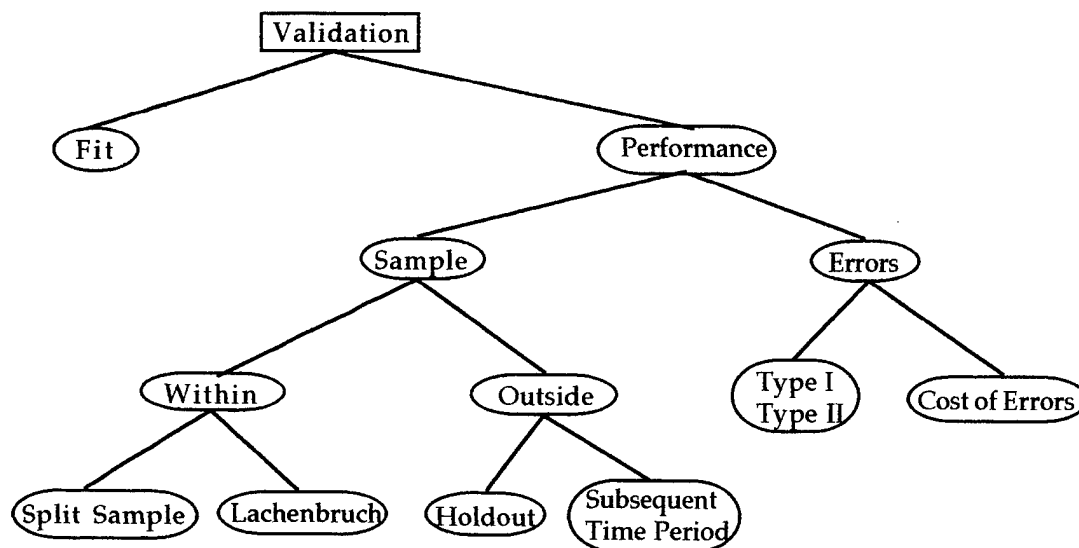


Figure 5. Model Validation Considerations

1. Fit of the Model

The quality of a model's initial results should first be evaluated using the statistical measures appropriate for the technique. The developer is seeking to understand if the statistical relationships are valid and if the independent variables capture adequately the characteristics of the dependent variable. This is the ex post examination of the quality of the variables: whether they have sufficiently discriminated between the failed and non-failed businesses in a manner statistically different from a chance distribution. As the R^2 , F-statistics, t-tests, and Durbin-Watson statistics are indicators of overall model significance in linear regression, each of the above modeling techniques has its own measures of significance which should be applied by the model's author and reported. The fit also relates to how well the model classified the development data.

2. Performance of the Model

Once the model has been developed and the fit of the data is deemed statistically significant, the next validation step is to assess the performance of the model in use. There are two ways to validate the performance of the model. First, is how well it performs on a sample of data distinct from the one used for the model development. These issues will be

discussed first. Second, is an analysis of the errors generated by the model, analysis of the costs of those errors, and what can be done to minimize those costs.

a. Sample

In selecting a sample on which to apply the model to assess its performance, the developer has a fundamental choice. The model may be tested on a sample taken from within the development data or from an entirely new set of data. Recalling the discussion on sampling, there is a tension between sample size and relevance. The more relevant a model intends to be for a particular population, the smaller the available sample of data. This problem was discussed in the context of the development sample, but it is equally applicable to the validation sample. Due the scarcity of data on failed firms, the use of a validation sample taken from within the development sample may be necessary for purely practical reasons. There are two common options available to the researcher if the choice is made to use within sample data.

The first choice is the use of a split sample. Here, the model is developed using all of the data and then tested with some subset of that data. The second choice is the Lachenbruch technique. This technique uses the development data by reconstructing the model using a sample containing $n-1$ observations and then predicting the missing observation. This is repeated n times. The summed classification error rate of the n models becomes the validation error rate. It is a useful procedure when dealing with a small sample size.

If the developer has the luxury of a larger population or a longer period of time (and the construct of the research permits it), the use of an outside sample is the preferred method of validation. There are two common options facing the researcher here: a holdout sample from the same period of time as the development sample, and a second sample from a subsequent time period.

When using a holdout sample, the original sample is divided randomly into two groups, one group is used to develop the model, the second group is used to validate the model. The principle criticism of the use of a hold-out sample is that the homogeneity of the data yields validation results which are biased upward.

To alleviate this criticism, validation data can be taken from a time period subsequent to that of the development data. The original data set is split chronologically, the earlier data is used to develop the model, the latter data to validate the model. This is particularly important if the model is designed to predict a future failure event, rather than discriminate between companies which may have already failed or not. Ideally, the model should use data from time t to assess the risk of failure in time $t+1$. Then the model should be evaluated using data from time $t+x$ ($x \geq 1$) to assess the risk of failure in time $t+x+1$. This issue is discussed in detail in Joy and Tollefson (1975).

b. Error Types

There are two types of errors a model will exhibit. A Type I error occurs when the model assigns a business to the non-failed category when in fact it did fail. A Type

If error occurs when the model assigns a business to the failed category when in fact it did not fail. The error rates should be determined and reported for the original development data for the model and then recomputed using one of the aforementioned validation techniques. The goal of the model, naturally, is to minimize the error rates. But to ensure the model is performing most efficiently, the user must consider the cost associated with each type of error.

c. Costs of Errors

The costs of errors refer to the economic costs associated with the model providing misleading information. It may be most efficient to have no errors of one type at the expense of a higher error rate of the other type. The classic example is medical testing: it is much less costly to receive a few more "false positives" when testing for the presence of disease and risk raising patient anxiety levels, than to have "false negatives" and needlessly risk the patient's life by failing to detect the disease.

Eisenbeis (1977); Dopuch, Holthausen, and Leftwich (1987); and Koh (1991) provide nearly identical models for expressing the costs of errors:

$$EC = (P_I * P(F) * C_I) + (P_{II} * P(NF) * C_{II}) \quad \text{Eq. 3}$$

In short, the model states that the cost of errors (EC) is equal to the sum of the cost of each type of error. The costs of each type are expressed as the product of the probability of committing that type of error (e.g., P_I is the probability of committing a Type I error) times the probability that the firm belongs in the other category ($P(F)$ and $P(NF)$) times the cost of that type of misclassification (C_I and C_{II}).

The critical value, be it a discrete value or probability, should be determined such that it minimizes the costs of errors. Depending upon the use for the model, the costs of these two types of errors may be very different. For instance, a commercial bank using a model to detect loan default risk at a time when business is plentiful would find it more costly to provide a loan to a business that eventually fails (Type I error), than to deny a loan to a business which may in fact be perfectly capable of repaying it (Type II). On the other hand, a government user may erroneously provide support to a vital industry whose key businesses appear unhealthy, but actually aren't (Type II) at large expense to the taxpayers. A consideration of these costs of errors must be made by the user to ensure that the model is not only accurate in an absolute sense, but that it is also providing economically efficient results.

The user of the model must also assess the context of the application of the model in determining the critical score or cut off point. The critical value which best discriminates between businesses that have already failed may differ from the value which best discriminates between businesses that will fail in the future. Differences may be caused by the effect of time on the interrelationships between the variables, the difference becoming

more pronounced the further into the future the model intends to predict. In fact, there may be more than one critical value to distinguish between failure in one year versus failure in two or three years. The determination of that critical value is a vital issue to the user.

G. SUMMARY

This chapter has provided a framework for the analysis of financial scoring models. Six dimensions related to constructing financial scoring models have been identified, explained, and the choices and issues surrounding each have been discussed: the theoretical basis for the model, the sample selection and data collection process, the dependent variable and definition of failure, the independent predictor variables, the modeling techniques, and the validation process. The next chapter will critically evaluate the literature along the same dimensions, the goal being a snapshot of the state of the art of the use of financial scoring models used to predict business failure. Chapter V will then use this framework to make recommendations for improving failure prediction in a DoD context.

IV. THE STATE OF THE ART

Last chapter, a comprehensive framework was developed for analyzing financial scoring models. Six dimensions related to model construction were identified, and issues were introduced which must be considered by both the developer and user of the model. These dimensions are equally useful for evaluating the related literature. The author has uncovered 33 different works which have developed financial scoring models and scores of other works which have impacted model development and address issues relevant to the field. In this chapter, those models and the related literature will be evaluated along the six dimensions to provide an assessment of the state of the art of the use of financial scoring models to predict business failure. While it would be possible to evaluate all models along all dimensions, it would not be very efficient. The literature will be discussed as contributions are made to the state of the art; some models may be mentioned only once, others several times.

A. THEORETICAL BASIS FOR MODEL

Within the literature there are two areas of theoretical work which relate to the use of financial scoring models to predict business failure. The first addresses the behavior of the firm and how the activities of the firm are related to failure: theories regarding liquidity and the need to maintain sufficient cash flows to avoid failure, the events which may determine a firm's ability to survive, and the actions taken by firms in periods of financial distress. The second area of theoretical work addresses the content of various information sets used as predictor variables. The literature reviewed here is principally concerned with the nature of financial ratios as applied to failure prediction, and, to a lesser degree, the applicability of literature from the field of auditing.

1. Theory Regarding the Behavior of the Firm

There exists no universally accepted theory of business failure, but that has not prevented the application of theory to the task of predicting failure. While many models have been constructed without regard to theory – using merely statistical techniques to show the existence of some relationship between failure and a set of predictors – others have relied upon theory to guide the development of a model. But this has only become prevalent in recent years. Zavgren (1983) concluded in her assessment of the state of the art, "An analysis of the literature indicates that considerable progress has yet to be made in both understanding the causes of financial distress and in acquiring the ability to predict it." As this chapter unfolds, it will become evident that progress has been made on both fronts.

a. Cash Flow Theories

The Issue: of what predictive ability is an analysis of a firm's cash flow?

Cash flow theories, also referred to as liquidity theories, assert that the failure of a firm is directly related to the status of its cash balances. Through normal business operations,

financing arrangements, and investments, a firm generates a stream of incoming cash. This stream of cash is offset by the expenses incurred in the same three activities: operations, financing, and investing. At the beginning of a given period, the firm has a supply of cash on hand which is either added to or subtracted from as a result of the inflows and outflows of cash described above. At the point at which the supply of cash is depleted, the firm is forced to default on obligations, and is said to fail.

The Literature. Until Beaver (1966), little had been written on the subject of failure prediction. Beaver's seminal work introduces the cash flow theory of firm failure, a theme often repeated in the literature. He used this theory to identify 30 potentially useful financial ratios and tested the predictive ability of the financial ratios using a univariate discriminant method. He found six to be significant predictors of failure. Blum (1974) developed several multivariate discriminant models based upon a cash flow theory which influenced the choice to use measures of liquidity, profitability, and variability as independent variables. Van Frederikslust (1978) theorized that failure occurs "at a certain moment when its cash flow from operations plus the proceeds from new loans and liquidation of assets plus the opening balance of liquid reserves is insufficient to pay the obligations due for that moment." But he found estimating the amount of cash needed at the time of failure very problematic and data insufficient to construct a predictive equation. Eventually, he extended the information set of independent variables and built a model using those variables which were statistically significant, ignoring his underlying theory.

Wilcox (1979) applied a "gambler's ruin" theory to the practice of predicting business failure. Gambler's ruin theory states that a gambler begins with a certain cash balance, and in a series of successive trials, has a probability of increasing the holdings by one dollar equal to p , and a probability of losing one dollar equal to $1-p$. The game continues until the gambler has run out of money. Wilcox replaced the gambler with the firm, and failure was defined as the moment when net worth equaled zero. The problem he encountered was that the events which drove the increases and decreases in net worth needed to be assumed and probability estimates applied. The cumulative probabilities which resulted were so uncertain and unreliable, he abandoned the model.

John and John (1992) and John (1993) both use a theory of illiquidity to define financial distress; illiquidity being "a mismatch between the currently available liquid assets of a firm and its current obligations under hard financial contracts." They state that as illiquidity equates to financial distress to emerge from the distress the firm must liquidate assets or convert the hard contracts to soft ones. Examples provided for hard contracts include coupon bond payments; soft contracts are common and preferred stock payouts. They showed that firms in specialized industries are particularly vulnerable to financial distress. Their proxies for specialized industries were the level of advertising expenses and research and development expenses. Considering the high amounts of research and development spending among DoD contractors, the theory suggests they may be particularly vulnerable.

Platt (1995) also used a cash flow model to base the selection of independent variables for his study of failures among firms which had recently transitioned to public ownership.

What we know. The cash flow theory of failure had widespread influence on the development of failure prediction models in the early years of the research, but was nearly abandoned during the 1980s. It is interesting to see the theory revisited in recent years by John, John, and Platt. While the cash flow theory was not widely used in the 1980s, two other themes emerged which took a more holistic view of the firm. First was the abandonment of theory for purely statistical and mathematical models which looked at all aspects of financial health. Jones (1987) concluded in his assessment of the state of the art, "Overall, most bankruptcy researchers have not applied theoretical models to empirical research...the more sophisticated models have been based on statistical or mathematical literature and have not provided economic guidelines to aid in independent variable selection." The second, was the emergence of an events approach to failure which considers qualitative as well as quantitative elements. The latter will be discussed next.

b. Theories Regarding Failure Events and the Actions of Distressed Firms

The Issue: of what predictive ability is an examination of the events associated with failure? A second set of theories regarding the failure of firms is related to the events which precede failure and the actions taken by firms that fail. A postmortem examination of failed firms may provide insight to events or conditions which may be used predictively in a scoring model. This postmortem study will be enhanced by also studying the actions taken by distressed firms that do not fail. How do these two types of firms differ? What caused the failure of one and, conversely, what actions saved the other?

The Literature. Bulow and Shoven (1978) examined the conditions that force a distressed firm into bankruptcy by analyzing three asymmetrical claimants: bond holders, banks, and equity holders. The authors believe that bank and equity holders could align and make uneconomic decisions regarding the actions of the firm at the expense of bondholders. Hudson (1986) studied why firms go insolvent using a time series analysis and discussed the differences in pressures exerted on the firm by trade creditors and banks. He found profits were significant in determining the number of liquidations. He also showed age to be highly significant – younger firms failing more frequently than older, established firms. Wruck (1990) also examined the effects of claimholders, finding that different claimholders will interpret the same information in different ways, maximizing their self-interest at the expense of the other claimholders, if necessary, and often with detrimental effect on the firm.

Schwartz (1982) examined financial reporting decisions made by firms facing increased risk of insolvency. He found that distressed firms made nearly twice as many material changes in financial reporting practices as healthy firms and over four times

as many material income-increasing changes. He defined material as increasing income by 25% or decreasing it by 15%.

Giroux and Wiggins (1984) developed an events approach to bankruptcy. They created a process model for evaluating the financial deterioration of declining firms. The premise was that poor performance leads to "policy and organizational changes to revitalize operations and maintain liquidity." They found that three events occurred in almost all cases of failure: net losses of income, debt accommodations, and loan default. Additionally, there was some evidence that dividend elimination and discontinued operations are predictive events, but not as significantly as the prior three. Outsider actions include: debt accommodations, credit restrictions, bond downgradings, and court actions.

Gilson, John, and Lang (1990) investigated the incentives of financially distressed firms to restructure their debt privately rather than through bankruptcy. Looking at 169 distressed firms in the period 1978-1987, they noted that half restructured their debt privately; those most likely to do so have more tangible assets, owe more of their debt to banks, and owe fewer distinct classes of lenders. John, Lang, and Netter (1992) studied 46 firms which voluntarily restructured in the 1980s in response to poor performance. They found no abnormally high turnover of top executives; a significant and rapid reduction of workforce or business segments; the ratios of cost of goods sold to sales and labor expense to sales declined rapidly; a cut in R&D expenditures; the occurrence of asset sales, dividend cuts, and increased investment; and a sharp reduction in the debt to asset level.

Ofek (1993) looked at 358 firms with a year of normal performance followed by a year of extremely poor performance. He found that higher predistress debt levels increase the probability of management action, particularly asset restructuring, dividend cuts, and layoffs. Opler and Titman (1994) found that firms with highly specialized products (as evidenced by high R&D costs) are especially vulnerable to financial distress.

Asquith, Gertner, and Scharfstein (1994) analyzed the avoidance of bankruptcy among distressed firms finding that debt structures which are complex (secured debt and numerous public debt issues) are impediments to restructuring outside Chapter 11 protection. Further, the ability to sell assets is affected by distress and leverage among other firms in the industry.

What we know. What has been learned from these studies is an appreciation for the issues which affect the firm beyond the obvious cash flow dimension of financial distress. It has been shown that other events may precipitate cash flow problems, may be by-products of it, or may represent early actions by the firm's management to fend off possible failure. Or, to put it another way, distress or failure tends to be preceded by or associated with a number of other observable events. Events of particular importance include losses of income, changes to accounting policy choices, debt accommodations, changes to dividend policies, and asset restructuring. (For some users of

a model, these events may be tantamount to failure.) Firms most likely to see these events associated with failure normally have complex debt structures and are in highly specialized businesses. This literature may serve to provide new indicators of distress, or distress avoidance, and may be of value when used in a failure prediction model.

2. Theories Regarding the Content of Information Sets

There are two sets of theories most applicable to the content of information sets from which independent variables used in financial scoring models are developed. First, are those theories related to financial ratios; financial ratios are clearly the most often used predictors of failure and this aspect of the literature is particularly relevant. The second set of theories relates to the growing literature on auditing and the usefulness of auditor's mental models used for making the going-concern judgment and how they can be applied to a financial scoring model.

a. Theories Regarding Financial Ratios

The Issue: what is the best way to employ financial ratios to predict failure?

As discussed in Chapter III, the use of financial ratios has strong appeal in predicting firm failure: they are reliable, obtainable, and intuitive. The use of a well-chosen set of ratios should capture the essence of the financial condition of a firm and be of some value in predicting its future condition. Chen and Shimerda (1981) wrote "the set of financial ratios used...should be selected in such a way that the ratios capture most of the common information contained in their factors and, as a group, contain more of the unique information than any other set of ratios." With this idea in mind, an extensive literature has evolved which has explored the nature of financial ratios and how they can best be arranged or categorized into a taxonomy which most efficiently describes the firm. In short, the process involves the collection of data from a large sample of firms, and the computation of as many financial ratios as practical; some studies have considered over 100 ratios. Factor analysis is used to group the ratios into categories in which ratios are highly correlated within the categories but not between them. Taken together, the groups should comprehensively describe the financial condition of the firms in the sample. The selection of one appropriate ratio per category, or factor, will minimize correlation between the ratios while maximizing their descriptive ability.

The Literature. Pinches, Mingo, and Carruthers (1973) developed seven empirically based classifications of financial ratios using factor analysis on 46 ratios. The classifications were: Return on Investment, Capital Intensiveness, Inventory Intensiveness, Financial Leverage, Receivables Intensiveness, Short Term Liquidity, and Cash Position. They also cited the seven ratios, one per category, which were most representative of the group.

Chen and Shimerda (1981) looked at 26 predictive studies that used financial ratios as independent variables. Of a total of 65 ratios cited, the authors used the 39 cited in the failure prediction literature as most useful and significant. They researched

those ratios and factors from previous studies concluding that seven factors (which mirrored Pinches, Mingo, and Carruthers) best described the financial condition of a firm.

Platt (1985) argued that financial ratios can be grouped into six categories: liquidity, debt, activity, profitability, growth, and value, and that the first four of these are relevant to bankruptcy prediction. Without citing a rationale, he provided, from within these categories, the "best bankruptcy-detecting financial ratios."

Moses (1995), in a manner similar to Pinches, Mingo and Carruthers and Chen and Shimerda, also looked at the categorization of financial ratios; his contribution was that the ratios considered derived solely from defense industry firm data. Beginning with 51 ratios from 50 defense contractors over the period 1983-1992, he found eight relevant factors. Of the eight, "three reflect the intensity or success of operations (turnover, profitability, and cash flow), [and] five reflect aspects of financial position (cash position, inventory, asset composition, liquidity, and leverage)." Of particular note is that these factors were found to be robust between industry segments, across macroeconomic cycles (both defense build-up and downsizing), and over time. For the DoD user, these findings are significant in that a comprehensive description of a defense contractor's health can be obtained by looking at eight common ratios.

What we know. The literature has demonstrated that the financial condition of a firm can be comprehensively described by as few as seven or eight common financial ratios, one from each factor derived. The processes used to generate these factors simultaneously ensure that they most comprehensively describe the financial condition of the firm while ensuring that they minimize the possibility of multicollinearity when used in a statistical model. Recently, these factors have been found to be robust across industry segments, macroeconomic cycles, and time.

b. Theories Derived from the Auditing Literature

The Issue: can the mental models used by auditors in forming a going concern opinion be incorporated into a financial scoring model to predict business failure? A basic tenet of auditing and accounting is that the firm is a going concern. This presumption is necessary for the accrual basis of accounting: assets are depreciated over time, for example, because of the expectation that the firm will continue as a going concern. An auditor has a responsibility under generally accepted auditing standards to alert the users of financial statements if, in the auditor's opinion, there are doubts regarding the firm's ability to continue. There exists an extensive literature around this point which addresses several issues; the issue of relevance here concerns the items an auditor finds of predictive value and which may be useful in the construction of a financial scoring model. Altman and McGough (1974) assessed the ability of models and auditors to predict firm failure, suggesting the usefulness of adapting an auditor's model to a mathematical one.

The Literature. Campisi and Trotman (1985) gathered the information auditors find most useful when rendering an audit opinion with an explanatory paragraph noting substantial doubt regarding the firm's ability to continue as a going concern. Based

on studies of auditors' decision making processes, they isolated those factors found to be most important and reliable to the participant auditors. In decreasing order, those factors are: cash position, short-term borrowing, retained cash flow, history of operating losses, deficiency in shareholders' funds, company history and industry information, and financial leverage.

Cormier, Magnan, and Morard (1995) used a risk analysis framework suggested by the auditing literature. This framework suggests that those indicators of inherent risk from the financial statements, as well as contextual qualitative factors which influence inherent risk and management motivation, can be used as predictors of firm failure. Inherent risk is the risk that there are undetected financial problems; from an auditing standpoint, it is that risk – risk that there is misrepresented information in the financial statements – that the auditor is trying to isolate.

What we know. While the literature is sparse, the concept is intuitively appealing: experts (auditors) who are charged with assessing the viability of firms have knowledge potentially useful to the art of predicting firm failure. Understanding the quantitative and qualitative factors auditors use to assess the financial condition of client firms may be useful in constructing a financial scoring model for predicting firm failure. Clearly, more work needs to be done in this area to obtain a widely-accepted set of factors used by auditors in a form that is useful to the construction of models.

B. SAMPLE SELECTION AND DATA COLLECTION

As introduced last chapter, there are two categories of issues related to the selection of a sample and the collection of data. The first are the conceptual issues which relate to the industry from which the firms derive, the economic and business climates in which they operate, and the timing of the data with respect to the failure event. The second set of issues are practical ones: the availability of relevant data, and the nature of the composition of failed and non-failed firms in the sample.

1. Conceptual Issues

Many of the works in the literature have failed to address the conceptual issues surrounding their samples. The most naive approach is to simply select all failed firms for which data is available during a relevant period of time. The next degree of consideration is to limit the firms to some industry segment and time period. This is as sophisticated as much of the research on failure prediction has been with respect to the sample. Altman (1968), the most often cited model, simply chose those industrial firms that filed for bankruptcy between 1946 and 1965. Others, however, have paid particular attention to the sample and made it the point of their contribution to the field.

a. Firm Size and Industry

The Issue: how have the bounds of firm size and industry affected the composition of samples used in developing financial scoring models? The developer of a

model must decide how broadly applicable the model will be. This can be done, in part, by limiting the sample along the dimensions of the size of the firm and the industry in which it operates. The distinction of a particular industry may be relevant to those studying that industry, of course, but may also have other ramifications. For instance, as Chapter III introduced, a firm in the financial services industry will have very different values for a given financial ratio than a manufacturing firm will, yet they may both be equally healthy within their given context. This subsection will look at the firm size and industry contexts which have been isolated and studied within the literature.

The Literature. Edmister (1972) and Keasey and Watson (1986) used small businesses as their sample, a particularly problematic group due to data availability problems. Edmister developed a failure prediction model using data obtained from the Small Business Administration. His results were comparable to those of prior studies conducted on larger firms. Keasey and Watson developed a linear discriminant model for predicting small company failures. Their model performed significantly worse than the majority of large company studies. They supposed that it may have been due to the unreliability of small company data. Moses and Liao (1987) also studied small firms, concentrating on private government contractors. Platt (1995) studied the failure of companies that had recently held an initial public offering (IPO), that is, transitioned from private to public ownership. While not all IPOs are small firms, they are unseasoned and the data available is often scarce and potentially unreliable (unaudited).

Nearly all of the models in the literature limit the industry being studied to either industrial firms or banks and savings and loans. This is a necessary distinction because of the differences in regulatory influence and financial structure. As this thesis is intended for a DoD audience, it will focus on those models which have considered industrial firms. Of those studies that looked solely at industrial firms, very few limited their focus to a specific industry segment. Mensah (1984) and Schary (1991) both isolated industry segments and are described in the next subsection of this chapter since their principal contribution was related to the economic context. Among the other works to isolate industry segments, defense firms have been very common.

Matthews (1983) used a purely qualitative set of variables (the complexity of the language used in annual reports) to predict failure among defense contractors. Moses and Liao (1987) used a sample of small, privately held government contractors to develop their index model. Dagel and Pepper (1990) developed a model using firms which represented DoD contractors (about one third were actual contractors, others were engaged in similar businesses). Christensen and Godfrey (1991) also limited their sample to DoD contractors, using both logit regression and discriminant analysis to build models.

What we know. Of the 33 studies reviewed by the author in which models were developed to predict failure of firms, only four focused specifically on small firms and six considered the issue of differences between industries. The field has acquired only limited, and sometimes conflicting, knowledge. For instance, Mensah (1984) concluded

that different models are appropriate when addressing different industry segments, but Moses (1995) demonstrated in his study of the nature of financial ratios that there is robustness across industry segments. Small businesses have not received much study, mainly due to data availability problems which will be addressed in Subsection 2 below. At this point, it is unclear whether a model must consider, or be limited by firm size and industry segment to be useful; further research must be done in both areas to adequately answer this question. For the DoD user, it is encouraging to see the defense industry segment isolated, but as Section F of this chapter will show, the models developed thus far require further refinement.

b. Business and Economic Climate

The Issue: how have considerations of the business and economic climate affected the composition of samples used in developing financial scoring models? In addition to considering the size and industry of a firm, the other conceptual issue that faces the field relates to the business and economic climates. The issues raised are, first, how the climate of the industry or economy affects the performance of the model; that is, whether the model will be validly applied during economic expansionary, recessionary, and stagnant periods. The second issue involves the study of already distressed firms. It can be argued that the challenge of building a failure prediction model is not the discrimination of failed firms from healthy ones, rather it is the discrimination of failed firms from non-failed, but otherwise financially distressed, firms. Jones (1987) raised this point, "It may be assumed that the decision-maker can distinguish between fairly healthy firms and firms in very serious financial distress. A real test of usefulness of the model is its ability to distinguish between marginal firms." The following set of works from the literature attempts to do just that.

The Literature. Dickerson and Kawaja (1967) showed a correlation between the rate of business failures and the business cycle. This was confirmed by Rose, Andrews, and Giroux (1982) who developed a model to predict the failure rate of businesses using macroeconomic data as independent variables. Many model authors have recognized the effects of macroeconomic conditions, but rather than address it explicitly in their models, they chose to pair failed and non-failed businesses in the sample to minimize the effects. While matching the sample will hide the effects of the macroeconomic influences (since they will presumably affect all firms in the sample equally), recognizing the effect and then minimizing it adds nothing to the body of knowledge regarding the predictive nature of macroeconomic conditions. (This practice of matching introduces new problems which will be discussed in Subsection 2, the Practical Issues.)

The sole model uncovered by the author that specifically isolated macroeconomic conditions was Mensah (1984). Bridging the research between industry segments and economic climates, Mensah (1984) isolated both. Using the same data set in all applications, he found that different models are appropriate when addressing different industry segments, and that the accuracy of the models differed across economic climates.

In considering the specific firm's business climate, several studies have been conducted. Schary (1991) looked at an industry in decline (textiles in the 1920s to 1940s) with the goal of predicting the form of exit of a firm: merger, voluntary liquidation, or bankruptcy. Her contribution to the issue of samples is that she, first, limited the sample to an industry in decline, and, second, did not limit that sample to a single category of exit. Her work suggests that other research may be overly sample specific since financial distress can be manifest in ways other than bankruptcy that may be relevant to the user. (Many other works have focused on businesses in decline and were introduced in Section A.1.b. of this chapter. Those works include: Bulow and Shoven (1978), Hudson (1986), and Wruck (1990) studies on claimholder incentives and actions; Giroux and Wiggins (1984) study on events approaches to failure; Gilson, John, and Lang (1990) and John, Lang, and Netter (1992) studies of voluntary restructurings; Ofek (1993) and Opler and Titman (1994) studies related to the effects of capital structure on response to distress; and the Asquith, Gertner, and Scharfstein (1994) study on the avoidance of bankruptcy by junk bond issuers.)

What we know. It has been shown in the literature, and makes intuitive sense, that the macroeconomic environment affects the rate at which firms fail. This issue is relatively unambiguous and, in only one instance, has it been isolated and confirmed with a failure prediction model applicable to individual firms. Other researchers have conceded the point and made efforts to minimize the macroeconomic effects on the model in an effort to isolate other discriminating factors.

Another lesson learned related to the sample is the issue raised by Schary: failure is manifest in many ways. Sample selection must be careful to recognize that fact and the author of a model must be certain to capture a sample which reflects failure in a way relevant to the user of the model. (Section C of this chapter will cover this issue in more detail.)

c. Sample Size vs. Relevance

The Issues: how has the body of literature responded to the inherent tension between sample size and relevance? In Chapter III, the tension between sample size and relevance was introduced and shown to be especially problematic in the study of firm failure. The extremely low rate of firm failure forces the researcher to either work with a sample so small the statistical significance of results are questionable, or to expand the sample by stretching the boundaries of time or industry, perhaps reducing the sample's relevance.

The Literature. The literature has responded to this issue primarily by accepting smaller samples. Few of the models' authors specifically mentioned the issue, but the problem is evident looking at the sample sizes used in the studies presented in Table 2. Among those who have addressed the problem is the Dagel and Pepper (1990) study. They wrote that "the degree to which a reasonable size sample of DoD hardware contractors could be assembled was limited by the number of recent bankruptcy cases." Aly, Barlow,

Author	Year	Failed Firms in Sample
Altman	1968	33
Dambolena & Khoury	1980	23
Zavgren	1985	45
Moses & Liao	1987	29
Dagel & Pepper	1990	29
Seaman, Young, & Baldwin	1990	41
Aly, Barlow, & Jones	1992	26
Platt	1995	32

Table 2. Sample Size of Selected Failure Prediction Models

and Jones (1992) found that to get a relevant sample (specified time frame, public firm, industrial, and proper data available) they had to consider the entire population that met the criteria. Several other studies have had the same restriction. The problem is that it leaves no firms available for validation and the results are likely to be sample specific.

What we know. There is an inevitable tradeoff between sample size and relevance. Research to date has acknowledged the problem, but found no way around it. Thus, models are forced to balance concerns of internal validity (improved by larger sample sizes) and external validity (improved by a well-bounded, relevant, and thus, smaller, sample).

Compiling a relevant and large sample is, and, for the foreseeable future should continue to be, a problem. It is expected that as data becomes more accessible and less expensive to obtain, the problem will be alleviated somewhat. But the amount of data is still limited by the extremely low rate of business failure; and there is no reason to expect the failure rate to rise.

2. Practical Issues

Chapter III discussed two practical issues related to the sample: the composition of the sample (across the categories of failed and non-failed) and the availability of data.

a. Composition of the Sample

The Issue: what are the effects on a model from using other than a matched sample of failed and nonfailed firms? The decision facing the developer of a model related to the composition is whether to use a matched-pair design (one failed firm matched with a similar non-failed firm) or to approximate the relative proportions of failed firms to non-failed firms in the population or to use some other combination.

The Literature. In considering the composition of the sample, approximately two-thirds of the models examined by the author used a matched pair design. In short, a set of failed firms is derived using the conceptual framework chosen by

the model's developer and as limited by the availability of data. That set of firms is then matched – normally by industry, date of financial data, and size of firm – with a nearly identical healthy firm. The model is then tasked with discriminating between the two sets. (In much of the literature describing the actions of distressed firms, the entire sample is comprised of distressed firms and there is no need to match them with healthy firms in any proportion.) The few works which have deviated from a matched pair design follow.

Ohlson (1980) was the only model uncovered that attempted to use prior probabilities as true to historical norms as possible. His sample consisted of 105 failed firms and 2058 randomly chosen non-failed firms. His model predictions, however, were not significantly different from a chance classification based upon the prior probabilities. Lau (1987) came close to matching prior probabilities; in constructing her five-state model, she considered 350 healthy firms and 20, 15, 10, and 5 firms for each of the other four declining states of financial health, respectively. Frydman, Altman, and Kao (1985) used a data set of 58 bankrupt firms and added 142 randomly selected non-bankrupt firms to generate a total sample size of 200. Coats and Fant (1993) used a 1:2 ratio of failed to non-failed firms in developing their neural network model. Cormier, Magnan, and Morard (1995) used a sample of 138 failed and 112 non-failed firms, and Platt (1995) used 32 failed and 76 non-failed; both of these studies used all available data that met their inclusion criteria.

Zmijewski (1984) discusses biases which may be introduced when using non-random samples. Specifically, he suggests that when samples include more distressed firms than in the general population, the model will be biased toward classifying a firm as distressed. While this bias exists, there is no impact on the statistical inferences which can be made by the research. Where the bias is felt is in application: if the model is biased toward returning a failed classification, then the Type II error rate will be higher. Depending on the modeling technique, this may be correctable by adjusting the cut-off score for classification. (Error rates will be discussed in more detail in Section F.)

What we know. Very little consideration has been given to the practical issue of sample composition. Most of the research has been conducted using a matched pair design; those studies using other proportions did not do so with the intention of specifically studying the effects of using other than a matched pair. Matching has been done mainly to minimize the effects of macroeconomic conditions, firm size, and other factors known to influence the failure rate, but which are not the purpose of the developer's research.

There is another issue the author has yet to see addressed in the literature regarding the use of a matched-pair design. When using a development sample matched so that the effects of macroeconomic conditions, firm size, and other factors, are minimized, the model better discriminates between the firms in the sample by isolating those characteristics that are unique to the firms in each category (failed and not failed). However, when the model is applied to a new sample (or a specific firm) to predict its

future state, the data entered into the model is not isolated from the same effects. The factors whose effects were minimized in development are now certain to be different, potentially reducing the discriminating power of the model. In other words, the model will assign coefficients to the independent variables which best discriminate between failed and non-failed firms given the context of the development data. When the context changes, as it will in application, the model is actually telling the user what the likelihood of the firm failing would be in the context of the model's development, not what the likelihood of failure is in the current context.

Zmijewski (1984) and the author have presented issues which require further study in order to be certain that the current practice of using a matched-pair sample, given its faults, is the preferred way to develop a useful model. As stated last chapter, if the intent of the model developer is to only test for relationships between certain independent variables and the failure event, then these issues are not germane. It is expected that research of that type be clearly identified as discovery and, because of the practical limitations, not be identified as a model to be used in practice.

b. Availability of Data

The Issue: is there a problem of data availability and how has the field responded? Data availability can also present a practical problem for the construction of a relevant data set. As stated last chapter, Zmijewski (1984) discusses the biases introduced when limiting a sample to only those firms for which a complete data set is available. Other issues which may affect the sample are filtering mechanisms imposed by the data source: the bases on which the source included or excluded data.

The Literature. There have been multiple problems cited in the literature with respect to data availability, the most common being a limitation on the number of firms with sufficient data. Christensen and Godfrey (1991) identified 150 government contractors who filed for bankruptcy, but could find sufficient information for only five of them. Aly, Barlow, and Jones (1992) sample was restricted by the lack of current cost data. John (1993) was frustrated by the availability of a specific data point (Tobin's Q, the ratio of market value to the replacement value of the firm's assets) during the relevant time period. And Platt's (1995) IPO study sample was reduced from a potential size of 135 failed IPOs to only the 32 for which complete data was available. Most other studies suffered a similar, but less severe, reduction in the potential sample due to data availability limitations.

Other works have experienced difficulties or raised issues related to data availability. Jones (1987) cautioned against using sources deriving their information from the news media (such as the *Wall Street Journal Index*). The issue he raised is that only data for newsworthy firms will be presented. Edmister (1972) has been criticized for using Small Business Administration (SBA) data for his small firm study because it limited the relevance of the model to only firms that applied for SBA loans. Moses (1990) noted an increasing number of analysts forecasts in more recent years, and a higher number of

forecasts for healthy firms than failing ones. This is consistent with other studies: as failing firms are often younger and smaller, there is often less data available about them.

What we know. The problem of data availability has affected the size of the samples used in some studies and has caused others to reexamine the generality of their models or the sufficiency of the variable sets. Availability is exaggerated by the fact that the failure rate for firms is small, and it becomes a greater problem when the source of the data may have also introduced some bias or limitation. The literature has coped with the issue in much the same way it has dealt with the issue of sample size versus relevance: the tendency has been to accept smaller samples. It is expected that, with the proliferation in recent years of easily accessed sources of data in electronic form, this practical issue will lessen in significance.

C. DEPENDENT VARIABLE

As introduced last chapter, the dependent variable raises both a conceptual and an operational issue. The conceptual issue relates to the question of the construct under study. The operational issues relate to the question of the scale used to measure the outcome. As some models are based upon a theory of firm behavior or movement into a particular state of financial health, these theories may affect the choice and composition of the dependent variable.

1. The Construct Being Investigated

The construct being investigated raises two distinct issues. The first issue to be discussed is the definition of failure being used in the construction of the model. The second issue is a question of timing: whether the model is designed to predict failure in n years, or if the model is designed to predict failure within the next n years.

a. The Definition of Failure

The issue: how has the field operationalized the definition of failure? The choice of a dependent variable goes to the construct being investigated, the definition of failure in use. The definition has a direct impact on other dimensions of the model such as the sample, the independent variables, and the modeling technique. Last chapter, the dangers of using the legal definition of bankruptcy were explored.

The Literature. The field began with Beaver's definition of failure including bankruptcy, bank overdrafts, or debt default, and Altman's definition of failure being the firm filing under Chapter X (now defunct) of the National Bankruptcy Act. Since these early studies, the definitions have become much more refined. Among those models using a cash flow theory, most have relied upon bankruptcy as the definition of failure, just as Altman had. But those applying the auditing theory have been more explicit.

Dopuch, Holthausen, and Leftwich (1987) and Koh and Killough (1990) defined their dependent variables as the rendering of a qualified report by an auditor. Cormier, Magnan, and Morard (1995) defined their dependent variable as a "potential

going concern,” operationally defined as annual stock returns of less than negative 50 percent. Coats and Fant (1993) used the rendering of a going concern opinion as their dependent variable “based on a desire to capture the ‘practical’ relevance of the predicted event.” Another reason cited by Coats and Fant mirrors Schary (1991): bankruptcy is only one possible outcome of financial distress. It can be argued, however, that since 43% of auditors fail to render a going concern opinion on firms that eventually failed (Menon and Schwartz, 1987), that the usefulness of this selection of a dependent variable is questionable.

Other theoretical bases for models have affected the nature of the dependent variable. Wilcox (1971) used a gambler’s ruin theory for the prediction of failure; his dependent variable therefore reflected the net worth of the firm and the eventual movement into the terminal state (failure). Edmister’s (1972) study of small businesses defined failure as default on the SBA loan. Schary (1991) and Lau (1987) each used dependent variables which took on values representing one of several states of financial distress. John (1993) used a modification of Beaver’s cash flow theory to create three models, each with a different dependent variable. The first was a liquidity ratio (cash plus marketable securities divided by total assets), the second and third were variations of debt ratios (short term debt divided by long term debt, and long term debt divided by total assets).

Bowlin (1994) studied the robustness of the operational definition of failure. He first tested preexisting models by Zavgren, Altman, and Dagel and Pepper in their original forms with new data. He then tested them again after relaxing the definition of failure. He found that “it does not appear that relaxing the definition of fiscal stress significantly alters a model’s prediction accuracy.”

What we know. The application of theory to the prediction of failure has resulted in an extension of the dependent variable beyond the common dichotomous failed/non-failed scheme of the early works in the field. That is, the field has recognized that failure has multiple meanings and has been operationalized differently by different researchers. While indicating a new level of sophistication in the literature, there is some unfortunate loss of comparability between works. Without a consistent metric, it becomes difficult to hold models up to scrutiny against each other. This trade-off, in the author’s opinion, is worth the additional knowledge gained in the field. The application of new theory will assist in developing new classes of independent variables and more insight into the dynamics of firm failure.

b. An Issue of Timing

The Issue: is the model designed to predict failure as of a specified point in time or within some range of time? Schary (1991) raises an issue yet to be addressed in the literature: the sample and model construction, are both affected by the timing of the predicted event. That is, the model can address one of two questions: what is the probability of the firm’s failure in n years, or, what is the probability of the firm’s failure within the next n years. This issue is particularly relevant to the user of the model. If the

author develops a model to predict failure in, say, three years, and a user finds his firm's data yields a score indicating it will not fail in three years, the possibility still exists the firm could fail in one year, or five years, or not at all. On the other hand, a model designed to predict failure within three years, may be too imprecise for some uses.

The Literature. To date, this issue has not been expressly explored. A minority of the models in the literature (Scharf included) have chosen to address the latter question, failure within n years (see Keasey and Watson (1986), Platt (1995), and Dagel and Pepper (1990)). Ohlson (1980) did both, generating three models with his data: one to predict failure in one year, one to predict failure in two years, and one to predict failure within the first two years. Clearly, most models have adopted the first question, predicting failure in a specific year. Many have developed models with the data pooled over time such that the information for the period immediately prior to failure is used enabling the model to predict failure in the subsequent period. (see Edmister (1972); Mensah (1983); Frydman, Altman, and Kao (1985); Moses and Liao (1987); Barniv and Raveh (1989); Seaman, Young and Baldwin (1990); Koh and Killough (1990); Koh (1991); and Goss, Whitten and Sundaraiyer (1991)). Still others have developed models using several years worth of data with the goal of predicting failure more than one year into the future. (Baldwin and Glezen (1992) looked 6 quarters ahead using quarterly data. Platt and Platt (1990) looked 2 years ahead. Lau (1987), Moses (1990), and Aly, Barlow, and Jones (1992) looked 3 years ahead. Altman (1968), Blum (1974), Dambolena and Khoury (1980), Matthews (1983), and Zavgren (1985) looked 5 years ahead. Rose and Giroux (1984) looked 7 years ahead.)

The last type of model results in one of two conditions: either the performance of the model diminishes as the failure event is further removed in time; or the research yields multiple models, a different one for predicting failure in each of several years. (The first issue will be discussed in Section F, Validation.) While the models of the second type normally provides some insight into the differences in the predictive ability of the independent variables across time, they are limited in application. A user would need to know, in advance, which model to choose. Applying data from the firm to all models may yield conflicting messages.

What we know. The literature has mostly provided models designed to predict the failure of a firm in a specified time period. This has provided precise models which accurately describe the development data and seem useful, but they can be problematic in application. The other technique, while used less often, is more useful in application, but can still provide mixed signals to a user. Section F of this chapter will discuss the performance of models and will explore the issue of usefulness in more detail.

2. The Scale of the Outcome

The Issue: *is the model capable of producing a discrete outcome or a continuous one, and if continuous, how is the distinction made between failed and nonfailed?* The

scale of the outcome can take the form of either a discrete or continuous measure. At one end of the continuum is a model designed to predict a unique event, e.g., bankruptcy, which would thus use a dichotomous dependent variable. Examples are the models using recursive partitioning, or artificial intelligence. Each of these provides for a dichotomous classification: either the firm is classified as failed or nonfailed. On the other end of the continuum is a variable that provides for a continuous outcome. Conditional probability models, discriminant analysis, and indexing are examples; their outcomes are numerical values that can be ordinaly ranked and lend a richer ability to compare the relative strengths of different firms or the same firm at different points in time. The conditional probability model differs from the other continuous outcomes in that it is in the form of a probability distribution, taking on values in the range of zero to one.

When using a model that produces a continuous outcome, the developer (or user) of the model must determine at what value the distinction will be made between failed and nonfailed. The option also exists to create a polytomous outcome by assigning multiple cutoffs representing varying degrees of financial distress. The selection of the actual value for the cutoff is addressed in detail in Section F, part 2.b., The Costs of Errors. For now, one simply needs to recognize that a cutoff score must be assigned.

The Literature. A continuous outcome was used by Altman (1968) whose multivariate discriminant model was assigned cutoff scores to provide for three classifications: bankrupt, non-bankrupt, and a "zone of ignorance" where the result was imprecise. Edmister (1972) and Dagel and Pepper (1990) used similar three-category classifications of multivariate discriminant scores. Moses and Liao (1987) and Moses (1990) provided an indices which they used in a two classification scheme, but, like discriminant analysis, could be adapted to more than two classifications of firms. Among other users of continuous outcomes, Ohlson (1980) pioneered the use of probabilistic outcomes in a failure prediction application. The use of a probabilistic outcome has become rather popular in the literature as evidenced by their use in the following studies: Zavgren (1985), Lau (1987), Barniv and Raveh (1989), Platt and Platt (1990), Koh (1991), Aly, Barlow, and Jones (1992), Cormier, Magnan, and Morard (1995), and Platt (1995). Lau (1987) developed her probabilistic model in such a way that cutoff scores provided for five different classifications; however, after validation of the model, it was determined that only two classifications were statistically different.

The use of dichotomous outcomes has been much less frequent. Frydman, Altman, and Kao (1985) introduced the use of recursive partitioning to the field, a technique that provides for a dichotomous classification. Coats and Fant (1993) built a model using artificial intelligence which also provides for only a dichotomous classification.

Some researchers have found it useful to build models that yield more than one measure of outcome. By applying different modeling techniques to the same set of data, they have provided both discrete and continuous outcome measures. This has been done, not for the benefit of obtaining the additional measure of output, but rather to test the

comparative accuracy of various modeling techniques. For this reason, research of this type will be discussed in Section E, below.

What we know. Models have normally been developed with a continuous outcome divided such that it yields two classifications, failed and nonfailed. Polytomous models, while less common, have also been developed. Until the introduction of logit regression analysis to the field (Ohlson, 1980), the only choice a user of financial scoring models had was a continuous outcome that could be assigned cutoff scores in an ad hoc manner to provide for two or more classifications. There now exists models which provide probability estimates, and most recently, others that yield a dichotomous outcome. What is most important, of course, regarding the scale of the outcome, is that the choice is purely a matter of the intended use of the model; the user should select the scale which best fits the application.

D. INDEPENDENT VARIABLES

"Ideally the researcher will draw on an economic theory in choosing those variables that will predict bankruptcy," wrote Jones (1987) in his assessment of the state of the art. That statement is still valid nearly a decade later and other conclusions drawn by Jones still apply, including: the state of the art employs accounting ratios as independent variables with few exceptions; many variable sets are based on a cash flow or liquidity theory; many researchers transform the variables to account for various macroeconomic effects; and the reduction of the variable set is based primarily on statistics and judgment, not theory.

What differs today from Jones' assessment is that, first, there is a greater emphasis on the trends inherent in the data and the stability of the data. Second, the boundaries of the information set have expanded to include more qualitative, capital market, and macroeconomic information. Third, theory is applied more often to the reduction and content of the variable set than ever before, particularly when the model is designed for application rather than discovering new relationships between failure and some set of predictors.

Chapter III developed the framework for analyzing the independent variables. Three broad issues were introduced related to the information set, the choice of a specific measure, and the criteria on which those measures are evaluated. Each of these three areas was developed in detail. The same framework is applied to the evaluation of the literature.

1. The Information Set

The information set relates to the question of the information content of the predictors of failure. What data, events, conditions, and actions are indicative of the failure event? The developer of the model has choices to make regarding the nature of the variables: qualitative or quantitative, firm specific or macroeconomic, accounting data or data from an independent source. This section will evaluate the failure prediction literature

along those tradeoffs and will discuss other literature which suggests predictive variables which have received little attention to date in failure prediction models.

a. Qualitative and Quantitative Variables

The Issues: how has the application of theory to financial scoring models affected the use of qualitative and quantitative variables? In what other circumstances has the literature found it useful, or suggested it would be useful, to use qualitative variables? Chapter III discussed the relative merits of both qualitative and quantitative variables. (In this discussion, a qualitative variable is synonymous with a dummy variable: information which is not readily converted to a numerical value but is incorporated in the model by using a (0,1) convention. Typically qualitative variables are created to reflect the presence or absence of some condition.) As most of the literature related to failure prediction derives from the field of accounting and finance, the use of quantitative variables is widespread. Until recently, it was unusual to see qualitative variables used in financial scoring models. The cash flow theory and the literature on the taxonomies of financial ratios suggest the use of quantitative variables, while the events approach to failure and the theory derived from the auditing literature suggest that qualitative variables are sufficient. It seems possible to develop a sound, theory-based model using one category of variables or the other. One could argue further that combining the two categories will yield a model with a richer informational content that could better capture the full picture of the firm's condition within its context. The literature, however, sends mixed signals.

The Literature. Lev and Sunder (1979) wrote that "the extensive use of financial ratios by both practitioners and researchers is often motivated by tradition and convenience rather than by careful methodological analysis." There are still some models being developed which rely solely on financial ratios due to their popularity in the literature (e.g., Koh and Killough, 1990; Dagel and Pepper, 1990; Goss, Whitten, and Sundaraiyer, 1991; and Baldwin and Glezen, 1992). In fact, over half of the models examined by the author used financial ratios exclusively.

As stated above, the use of a cash flow theory or the theory regarding the taxonomy of financial ratios suggests the use of quantitative variables. This is consistent with the independent variables actually used by the developers of models using these theories. Cash flow models were developed by Beaver (1966), Blum (1977), Lau (1987), and John (1993). Each used quantitative measures which are consistent with the cash flow theory: measures of cash flow, income, expenses, liquidity, and leverage. While Lau and John both used some qualitative variables, most were quantitative and consistent with the theory. Lau also considered the events approach to failure, (as evidenced by her dependent variable which measured five states of progressively declining financial health) which influenced the use of qualitative variables.

Several models used factor analysis – as in the research of Pinches, Mingo, and Carruthers (1973) and Chen and Shimerda (1981) – to guide the selection of financial ratios as independent variables. The influence of these theories is becoming more

common. Zavgren (1985) first used their taxonomies to guide the selection of the information set for her model. Others to use this theory include: Moses and Liao (1987); Platt and Platt (1990); and Aly, Barlow, and Jones (1992).

Other quantitative data used has included capital market information, macroeconomic indicators, data related to the industry segment, and analysts' earnings forecasts. These will be discussed in Subsections b and c.

Despite the widespread use of quantitative information, there is a strong argument to be made for the use of qualitative information. In fact, despite using financial ratios herself, Zavgren (1985) wrote:

Many unobservable factors influence the vulnerability of an individual firm. These include the unmeasured qualities of assets, the creative ability of management, random events and the decisions of regulators and courts of law. Any econometric model containing only financial statement information will not predict with certainty the failure or nonfailure of a firm.

Several models have incorporated qualitative data. Lau (1987) used several qualitative variables: the restrictiveness of the firm's loan agreements, the existence of a dividend payment, and the reduction or elimination of a dividend payment if previously made. The use of dividend payment cuts or eliminations is supported in the literature by Giroux and Wiggins (1984) events approach to bankruptcy; John, Lang, and Netter (1992) study of distressed firms that voluntarily restructured to avoid failure; and DeAngelo and DeAngelo (1990) study of dividend policies of distressed firms. John (1993) turned the failure prediction model around and predicted values for liquidity ratios using (among several quantitative measures) two qualitative variables, the Standard Industrial Classification code and whether or not the firm had filed for bankruptcy.

The most comprehensive use of qualitative variables is the model developed by Cormier, Magnan, and Morard (1995). Their model contained both quantitative and qualitative variables, the quantitative ones measured trends in the accounting data and will be discussed in the next subsection. The qualitative variables included: investment in new industries, a change in the number of the firm's operating locations (a sign of investment or asset liquidation), the implementation of bonus or profit sharing plans for employees, a change in the firm's controlling stakeholders, and changes in accounting methods. The last measure is supported by Schwartz (1982) who specifically studied the predictive value of changes to accounting methods.

Research influenced by the auditing literature and events approach to failure suggest the use of qualitative variables, variables that would capture the factors, events, and actions consistent with these theories. Four models reviewed by the author used these theories, but their incorporation of qualitative variables is not as expected. Koh (1991) and Schary (1991) did not use any qualitative variables, instead relying on financial ratios and measures of firm capacity. On the other hand, Lau (1987) and Cormier, Magnan, and Morard (1995) did use qualitative variables in their failure prediction models.

There are some qualitative measures which are suggested in related literature which have not been incorporated into failure prediction models. These include debt structure (Gilson, John, and Lang, 1990, and Asquith, Gertner, and Scharfstein, 1994), actions by claimholders (Wruck, 1990, and Bulow and Shoven, 1978), debt accommodations (Giroux and Wiggins, 1984), the age of the firm (Dickerson and Kawaja, 1967, and Hudson, 1986), and the rate of short term borrowing (Campisi and Trotman, 1985). All of these variables have been shown to be correlated with failure or the avoidance of failure in studies of financially distressed firms. It is surprising to see no use of them in the development of failure prediction models.

What we know. The literature shows a very strong bias toward the use of quantitative variables to comprise the information set used to predict failure. As the literature derives from the fields of accounting and finance, this is not surprising; there is little influence from the field of economics to suggest more qualitative measures. While the use of qualitative measures is still relatively sparse, there has been an increase in their use in recent years and it is expected to rise as the events approach to failure and the literature from auditing gain more acceptance. The most frequently used quantitative data are derived from the financial statements, and are normally in the form of financial ratios. The information content of those ratios usually centers around measures of profitability, liquidity, leverage, and cash flow management.

In short, the related literature suggests the use and the predictive ability of qualitative information, but it is an area yet to be adequately explored in the development of models. Recall from Chapter III, Hawkins (1986) asserted that much of a bond's rating is attributable to "management, industry, general economic conditions, future prospects, and other qualitative factors." In the next subsection, the economic and industry conditions will be explored.

b. Firm Specific or Macroeconomic Variables

The Issue: can failure be accurately predicted using only variables specific to the firm, or is it necessary to include variables measuring macroeconomic conditions?

The literature on the taxonomies of financial ratios suggests that a comprehensive description of a firm's financial health can be obtained by analyzing a few well-selected financial ratios. Some suggest this is sufficient to predict the future state of the firm. Other literature suggests a strong link between macroeconomic conditions and the failure of firms. The counter argument is that that information is already captured in the financial ratios and each firm is operating in the same economic context.

A second issue relates to the applicability of the model in use. As discussed previously, the model is developed using data from time t and will be used to predict an event in a later time $t+1$. The economic conditions will be different in time $t+1$, suggesting the usefulness of some measure of macroeconomic conditions in the information set of the model.

The Literature. Despite a literature that suggests the predictive ability of macroeconomic indicators – Rose, Andrews, and Giroux (1982) accurately predicted the rate of business failure using macroeconomic measures – their use in failure prediction models has been sparse and did not begin until 1990. Jones (1987) did not cite any models using macroeconomic variables and correctly concluded that “macroeconomic variables may be useful in forecasting, since it will be useful to predict the general probability of bankruptcy before assessing the likelihood of individual bankruptcy.”

Of all the models examined by the author, the first to use macroeconomic indicators was Platt and Platt (1990) whose model used industry average ratios. Cormier, Magnan, and Morard (1995) used investments in new industries and changes in the number of firm locations as indicators of industry health. Platt (1995) has been the only author to use widely available macroeconomic statistics, incorporating the prime lending rate and the percentage change in gross national product in his model.

What we know. Although a relationship between macroeconomic conditions and the rate of firm failure has been shown, the field is just beginning to incorporate macroeconomic indicators in failure prediction models. To be a useful tool, the model should contain a measure of macroeconomic conditions as both this author and Jones (1987) have indicated. In the discussion of the use of matched pair samples, the issue of timing and the differences between the conditions during the time period of the development sample and application sample were introduced. That argument is further evidence of the need for the use of macroeconomic indicators.

c. Accounting Data or Independent Analysis

The Issue: what are the comparative advantages of using variables derived from the firm's accounting data or data derived from some independent source of analysis?

One final issue related to the information content of the independent variables facing the developer of the model is whether to use data from the firm's accounting statements or to rely on an independent analysis of the firm. Accounting data has the appeal of being readily available and reliable, independent analysis has the added benefit of some level of interpretation of events by an “expert” who has considered both accounting and qualitative factors. Examples of independent analysis are capital market information such as stock prices and bond ratings, and forecasts made by financial analysts. These data may be used at face value or incorporated into some ratio such as dividend yield or book to market value.

The Literature. Like macroeconomic indicators and qualitative variables, the users of independent analysis are in the minority. Edmister (1972) was the first to use independent information. He generated intra-industry ratios using data from Robert Morris Associates and the Small Business Administration in his small company study. Blum (1974) was also a pioneer in the use of independent information. He computed the rate of return on stockholders equity and the fair market value of the net worth of the company using capital market information.

More recent users of capital market information are Rose and Giroux (1984) who considered numerous measures of stock price and performance in their original data set. Lau (1987) computed the trend in stock prices as an independent variable for her model. Dopuch, Holthausen, and Leftwich (1987) used four market variables to predict an auditor's going concern opinion: how long the firm's stock had been listed on a major stock exchange, change in the stock's beta, change in the residual standard deviation of returns, and common stock returns in excess of industry averages. Koh and Killough (1990) considered the ratio of market value to book value as a predictor. And after evaluating DoD specific models, Christensen and Godfrey (1991) recommended that future models incorporate market data to enhance their accuracy.

Other uses of independent analysis includes John's (1993) use of Tobin's Q ratio, the ratio of the market value of the firm to the replacement cost of the assets. This ratio was obtained from an independent source due to its computational complexity. A unique use of independent analysis is found in the model developed by Moses (1990). The only use of a forward-looking metric found by the author, the Moses model uses earnings forecasts by financial analysts. The benefit of this measure is that the analyst is presumably using quantitative, qualitative, firm specific and macroeconomic measures to reach a conclusion about the future prospects of the firm. As the model intends to predict a future event, this has strong appeal. While analysts' earnings forecasts have a reputation for being inaccurate (e.g., Dreman and Berry, 1995), this was accounted for in the model by the use of variables measuring error rates, biases, and the dispersion of forecast estimates.

Giroux and Wiggins (1984) showed a relationship between bond downgradings and failure. Bower and Garber (1994) also suggest the use of bond rating data and other capital market information. It is interesting to note that this measure has not been incorporated into models. Perhaps the low number of failed firms with publicly traded bond debt (refer to Table 1 in Chapter 1) is the cause.

What we know. While the literature, and common sense, suggest some usefulness of the data provided from independent analysts, the use of such data has been sparse in practice. This may be due to a perceived unreliability of the data, the confounding effects of biases introduced by the data source (discussed in some detail in Section B of this chapter), or a belief that accounting data is both necessary and sufficient. This perception is supported by the studies using factor analysis to create taxonomies of financial ratios and, perhaps more significantly, the accuracy of models based solely on financial ratios.

It is expected that, given the efficiency of the capital markets, data regarding stock price movement and fair market values will continue to be used. The use of other independent sources of information have added little to the state of the art. The use of other forward-looking measures is an area to be explored more fully, be they analysts' earnings forecasts, stock prices, bond ratings, or some other metric.

2. The Choice of a Specific Measure

Once the information set has been determined by the developer of the model, the decision turns to the specific measures used to represent that information. There are two broad issues to be discussed in this section: first, the construct and the selection of specific measures to represent that construct, and, second, transformations of those measures that may be appropriate.

a. The Construct and Selection of Specific Measures

The Issue: Is the selection of specific measures based upon some construct guiding the development of the model or a statistical reduction technique? In the first case, constructs will influence the choice of specific measures. For example, if the model is using a cash flow theory, the construct would dictate choosing measures which reflect information about cash flow such as liquidity, income, and retention of earnings. Narrowing the scope further, the researcher would then choose an appropriate measure to represent liquidity, income, and retention of earnings. Variables such as the current ratio (current assets divided by current liabilities), quick ratio (cash and marketable securities divided by current liabilities), net income, and return on stockholders' equity would be considered. The selection of specific measures should be based upon the theory employed and hypothesis tested by the model's developer.

Another issue related to the selection of a specific measure is the method employed to reduce the set of potential measures originally considered. Normally, a researcher will choose several measures to represent the construct. Through various techniques, the number of variables will be reduced such that the final model will include less variables than originally considered. The goal is to develop a model which retains a high level of accuracy with the minimum amount of independent variables. The reduction techniques are normally based in statistical relationships and are often accompanied by theoretical criteria and authors' judgment.

The Literature. Some of the models reviewed used the variables developed by previous authors in an attempt to illustrate the usefulness of a new modeling technique or to test them under different economic conditions. Examples include Deakin (1972) using the same 14 ratios used by Beaver (1966); Altman and McGough (1974) and Coats and Fant (1993) used the same variables as Altman (1968); and Barniv and Raveh (1989) used the same variable set as Frydman, Altman, and Kao (1985).

Baldwin and Glezen (1992) chose their measures based on timing of the information. Their hypothesis was that failure could be predicted sooner if the model incorporated quarterly financial data rather than annual data. Their results were inconclusive: there was no statistical difference in the predictive ability of the two sets of measures.

Some authors chose their variables based upon the hypothesis they were testing and retained all variables in the model, relying on their initial judgment to determine the optimum variable set. Moses (1990) used this procedure in his model employing

analysts' earnings forecasts. The forecast, various measures accounting for the accuracy of those forecasts, measures related to trends over time in the forecasts, and error measures were developed and employed in the final model. Goss, Whitten, and Sundaraiyer (1991) computed and used three ratios they "deemed to be predictors of bankruptcy:" the current ratio, the quick ratio, and an income ratio (net income divided by working capital). Cormier, Magnan, and Morard (1995) created 16 variables derived from the auditing theory they were testing. All 16 were used in the development of their discriminant and conditional probability models, but only nine were found to be statistically significant, *ex post*.

At the other end of the spectrum are models which begin with a set of variables and reduce the set based purely on statistical significance of the measures with respect to their ability to classify the development sample. The most common technique is to use a stepwise reduction method whereby the author specifies a threshold level of statistical significance the variable must meet and a computer program determines which variables are included in the model by evaluating each in turn based upon the prespecified performance criteria. Edmister (1972) was the first to rely solely on a statistical reduction and it has been used throughout the literature, most recently with Platt (1995). In fact, about one-third of the models reviewed used solely statistical techniques to reduce the variable set. Some model developers have been charged with what is pejoratively called "data mining" by considering large numbers of variables and allowing the computer to best fit a model by considering all of them and finding the best combination. The danger is that the model will "overfit" the development sample and lose relevance when applied to another sample. The most extreme case is the model developed by Rose and Giroux (1984) who considered 157 variables and reduced the set to 18 which were included in the final models.

An equally common technique is to temper the statistical reduction with a theoretical basis or the author's judgment. Those who used their judgment include Altman (1968) and Dagel and Pepper (1990). They were primarily concerned with ensuring the signs of the coefficients were appropriate; that is, the influence the individual variable had on the outcome of the model was logical. The other method involves the theoretical basis for the variable construct. For example, those who used a financial ratio taxonomy to influence the information set frequently select several ratios to reflect each factor, then apply a statistical reduction technique to reduce the field to (normally) one ratio per factor (provided *any* ratio for that factor was significantly significant). In doing this, the descriptive power of the taxonomy is preserved without introducing unnecessary correlation between the variables. Models of this type include Moses and Liao (1987), Platt and Platt (1990), Zavgren (1985), and Aly, Barlow, and Jones (1992). Similar approaches were used by authors employing other theoretical bases for their models, such as those influenced by the auditing literature (see Hudson, 1986; Koh, 1991; and Cormier, Magnan, and Morard, 1995).

What we know. Several techniques have been employed to select and reduce to appropriate levels the variables included in the models. The danger of overfitting is real and may have been violated by models employing only statistical techniques to reduce their variable sets. The models which have employed statistical techniques tempered with judgment or theoretical bases appear to be the better developed models. While ensuring the statistical significance of each variable in the final model, there is also an expectation that the set of variables did not lose its relevance as it underwent the reduction process and should be more generally applicable.

The categories of models first mentioned, those that use another author's variables and those employing no reduction technique, deserve specific mention. Those using another authors' variables have done so to introduce new modeling techniques (e.g., Deakin (1972), Barniv and Raveh (1989), and Coats and Fant (1993)) or new applications (Altman and McGough, 1974). Those that did not use a reduction technique are a mixed lot. If there was no reduction due to a well conceived set of variables (e.g., Moses (1990) and Cormier, Magnan, and Morard (1995)), this may be as useful as the models using a statistical reduction technique in conjunction with other criteria. However, models such as Goss, Whitten, and Sundaraiyer (1991) appear to be naively developed and one would be uncomfortable applying a model developed this way without rigorous validation.

b. Data and Variable Transformations

The Issue: in what circumstances has the literature found it useful to transform the data or the variables? Chapter III outlined various reasons why the data or specific variables may need to be transformed. The effects of time, changes to accounting principles, macroeconomic conditions, stability, and assumptions inherent in modeling techniques may call for the transformation of variables or data. Other transformations may occur because of the heightened predictive nature of a transformed variable over the raw data; for instance, the developer may find that trends in the value of certain ratios are more telling than the absolute values at a given moment in time.

The Literature. The literature shows that transformations of variables are becoming increasingly common, both to minimize the effects of some condition or to enhance the predictive ability of the variable. Transformations have included trend analysis, measuring the stability of ratios, computing averages over certain periods of time, industry-relative ratios, and other transformations to minimize effects external to the firm.

The first set of transformations are those categorized under trend analysis. Edmister (1972) first recognized the predictive value of trends in the data when he considered changes in inventory to sales ratios and the quick ratio. Blum (1974) used trends in net income and quick assets to inventory. Lau (1987) examined the trend in stock prices, capital expenditures, and working capital flow to predict failure. Dopuch, Holthausen, and Leftwich (1987) considered changes to the ratio of total liabilities to total assets, the ratio of receivables to total assets, and the ratio of inventory to total assets. They also considered changes in capital market data over time. Moses (1990) examined

within-year trends in analysts' earnings forecasts as well as year to year changes in forecasts themselves and the measures of error, bias, and dispersion. Platt and Platt (1990) and John (1993) included variables to reflect the rate of firm growth. Nearly all of the quantitative variables used by Cormier, Magnan, and Morard (1995) measured trends in profitability, working capital management, long term investment, and financial management.

Others have chosen to examine the predictive ability of the stability of variables and transformed the data accordingly. Blum (1974) used the standard deviations of net income and the ratio of quick assets to inventory in his model. Dambolena and Khoury (1980) computed for each firm, for each ratio, for each of five years prior to failure, four different measures of stability: the standard deviation over three years, the standard deviation over four years, the standard error of the estimate around a four year linear trend, and the coefficient of variation over four years. Developing models both with and without these measures of stability, they concluded that inclusion of the standard deviation greatly improved the accuracy of the model and that ratios for failed firms become less stable as the failure event approaches. Moses (1990) found that measures of bias, dispersion, and errors improved the predictive ability of his model using financial analysts' earnings estimates. Schary (1991) used the standard deviation of monthly return on equity and annual cash flow over five year periods. John (1993) used the volatility of operating income, defined as the standard deviation of earnings before interest and taxes divided by average total assets. (John (1993) also computed averages for the data. Most of the variables in the final model were averages of the information computed over the three years immediately preceding the failure event.)

Another common transformation, one designed to minimize the effects of macroeconomic and industry conditions, is the use of industry-relative ratios. Lev (1969) first suggested their use and Edmister (1972) first used the technique in a financial scoring model. He computed figures for firms relative to Robert Morris Associates small business averages and Small Business Administration averages. Lau (1987) also computed industry relative averages for the ratios of debt to equity and operating expense to sales. The use of industry-relative ratios has been championed in recent years by Platt and Platt (1990 and 1991) and Platt (1995). They have shown that ex post forecast accuracy is improved relative to ex ante forecasts when using industry-relative ratios.

Mensah (1983) examined the effects of inflation on predicting failure. He transformed the variables to reflect price level adjustments and current cost information. Unfortunately, he found that these adjustments do not greatly improve prediction accuracy.

What we know. As failure is a dynamic process, with the firm normally evolving over a period of years, the study of changes in the financial condition of a firm seems logical to the prediction of its failure. Therefore it is surprising to see that the study of trends in the data has been done less often than one would expect. Its frequency is increasing, however. Along the same lines, we have known that ratios tend to be less

stable for failing firms and industry-relative ratios are useful in isolating economic conditions and improving accuracy, but (until very recently) relatively few models have incorporated these transformations into their variable sets. The relative ease in computing these transformations in recent years may have contributed to their more frequent use. It is an encouraging sign, considering the effects these transformations have been shown to have on model accuracy. More study is needed, however, since very few of the variables shown to have predictive ability have been used in models in a transformed state.

3. Evaluation Criteria

Last chapter developed the framework for evaluating the quality of the variable set, both prior to model development and after. Prior to development, the aim is to evaluate the content of information represented by the variables. These *ex ante* criteria will ensure that the only variables considered are those with a reasonable expectation of contributing to the prediction of failure and that can be replicated in application to future samples. After development, the aim is to evaluate the variables' actual usefulness within the model. These *ex post* criteria ensure that the variable set used in the model sufficiently describe the condition of the firm and do so in a manner that makes rational and intuitive sense to a user and are generalizable across samples.

a. Ex Ante Criteria

The Issue: on what criteria should the model developer base an evaluation of the content of the independent variables? The criteria outlined previously for the *ex ante* evaluation are: obtainability, reliability, stability, and a basis in theory. Ideally, the variables will be based upon some theory of failure; barring that, there should still be some logical basis for their consideration. Regardless of the presence of a theoretical basis, the variable set should be obtainable with relative ease from a reliable source. Data that cannot be replicated by a user of the model is of little value and if the source or sources of data are unreliable, biases will be introduced and may possibly invalidate the output of the model. A stable measurement of the variables is also desirable: if the data are unstable in their application, the usefulness of the model again deteriorates.

The Literature. As most models use financial ratios or other accounting data taken directly from audited financial statements, most of the quality criteria are easily met. Certainly these data are obtainable and reliable; the stability of financial ratios could be questioned, however. Frydman, Altman, and Kao (1985) found it necessary to transform some variables due to changes to generally accepted accounting principles. Other transformations are conceivable due to differences in accounting policy choices such as depreciation and inventory valuation methods. The Aly, Barlow, and Jones (1992) study focused on these differences, examining the effect of transforming the historical costs in financial reports to current costs.

The criteria for evaluation are certainly a factor in the reluctance of researchers to use variables of a qualitative nature or those derived from independent

analysis. These types of data are frequently either difficult to obtain or of questionable reliability. Some examples from the literature follow. With respect to obtainability, Matthews (1983) required the use of an expert analyst to interpret the wording in the annual reports and John (1993) had difficulty obtaining a data point (Tobin's Q) for her sample. One must certainly question the reliability of independent analysis.

There is a strong tendency to use variables which were used elsewhere in the literature. Phrases justifying variable selection include "suggested by the literature," "advocated by prior literature," "frequently mentioned in the literature," and "significant in other studies." Users of such criteria for evaluating their variable set include Beaver (1966), Edmister (1972), Ohlson (1980), Dagel and Pepper (1990), Koh and Killough (1990) and Baldwin and Glezen (1992). This criteria can be taken to the extreme when a researcher replicates exactly the variable set used by another.

The final quality criteria is a basis in theory. Those authors who considered a theoretical basis for the selection of their variable set are shown in Table 3. Other authors did not consider an underlying theory in developing their variable sets, relying on the other criteria exclusively.

A	B	C	D
Cash Flow	Auditing	Taxonomy of Ratios	Events Approach
Beaver (1966)	Keasey & Watson (1986)	Zavgren (1985)	Lau (1987)
Wilcox (1971)	Dopuch, Holthausen	Moses & Liao (1987)	Schary (1991)
Blum (1974)	& Leftwich (1987)	Platt & Platt (1990)	
Lau (1987)	Koh (1991)	Aly, Barlow & Jones (1992)	
John (1993)	Cormier, Magnan,	Baldwin & Glezen (1992)	

Table 3. Theoretical Basis for Independent Variable Consideration

What we know. The popularity of accounting data and financial ratios as independent variables is due in large part to their quality *ex ante* and in part due to their popularity in the literature. The use of these ratios in conjunction with a theoretical basis for the selection of specific variables provides for even higher quality. While the use of financial ratios has been challenged in the literature, the challenges have been addressed, normally through transformation of the variable.

These quality criteria may also account for the reluctance to use qualitative and independent analysis variables in the model. They generally suffer from problems of obtainability or reliability despite their strong theoretical or logical appeal. The state of the art suggests that the first three criteria outweigh the significance of a theoretical basis and that attitude should continue so long as a widely accepted theory of failure eludes the field.

b. Ex Post Criteria

The Issue: on what criteria should the usefulness of the independent variables be evaluated? Once the model is developed – variable set is reduced to its final form, coefficients or cut-off scores are assigned, and outcomes determined – the variables remaining must be further evaluated along three additional criteria: their sufficiency, intuitiveness, and rationality. Sufficiency takes on a dual role: first, the variables must sufficiently describe the condition of the firm (and perhaps its context) such that they accurately discriminate failed from healthy firms; and, second, they must do so in such a way that there is not extraneous information that degrades the overall effectiveness of the model. Intuitiveness and rationality were described in some detail last chapter, but, in essence, ensure that the contribution each variable and its coefficient make are logical and consistent with underlying theory. These criteria can be assessed in two ways: through a statistical analysis appropriate to the modeling technique employed, and through the use of the author's own judgment.

The Literature. With the exception of two classes of models, all have evaluated the variable set after the model was developed using either statistical analysis or judgment. The two exceptions are, first, those models which chose the variable set in advance and employed no reduction technique thereby retaining all variables in the final model, and, second, those that used another researcher's variable set.

Statistical testing of the variables is a function of the modeling technique employed. Those researchers using discriminant analysis generally use F- statistics (e.g., Altman, 1968; Deakin, 1972; Altman and McGough, 1974; and Dagel and Pepper, 1990), and those using conditional probability models tend to use t- statistics (e.g., Ohlson, 1980; Rose and Giroux, 1984; Zavgren, 1985; Dopuch, Holthausen, and Leftwich (1987); Lau, 1987; Platt and Platt, 1990; Goss, Whitten, and Sundaraiyer, 1991; Schary, 1991; John, 1993; and Platt, 1995). Moses and Liao (1987) also used the t- statistic to test the significance of their index model. Frydman, Altman, and Kao (1985) used a cross validation technique¹ to test the variables used in their recursive partitioning model.

In cases where there are high levels of multicollinearity among the variables (common when using financial ratios), the contribution of individual variables is difficult to ascertain.² Some researchers have simply used judgment to assess the individual variables (i.e., it has the proper positive or negative sign) and only used statistical measure to evaluate the entire model (e.g., Mensah, 1983; Lau, 1987; and Schary, 1991).

¹ This is actually a validation test of the entire model, but since the modeling technique determines the variables to be included without permitting ex post modifications by the author, its use is tantamount to evaluating the quality of the variables.

² In an MDA model, the coefficients are not interpretable even in the absence of multicollinearity, only the ratios of the coefficients are unique. Thus, standard t- tests of their significance are inappropriate. There are several alternative approaches which go beyond the scope of this thesis; the reader is referred to Eisenbeis (1977) or Altman, et. al. (1981).

The use of factor analysis and stepwise reduction techniques to reduce the variable set contains implicit ex post quality criteria. In fact, as the basis for reduction of the variable set is a prescribed level of statistical significance, most of the ex post and ex ante criteria are met simultaneously, but not all of them. As stated earlier, when using the stepwise technique there exists the danger of overfitting and the introduction of irrational variables. From the review of the literature, there does not appear to be a problem with irrational variables appearing in the models, implying discretion by the model developer.

The researcher's judgment is also used to test for the intuitiveness of the variables. Most authors have examined the variables to assure the coefficients have the correct positive or negative sign. It may be possible, especially when there are high levels of correlation between the variables, to have counterintuitive signs on the coefficients. Nearly every work reviewed by the author cited some use of judgment on the researchers part to ensure ex post quality of the variable set.

What we know. The research reviewed has been diligent in testing the ex post quality of the independent variables. Both statistical and judgmental techniques are being used to ensure the sufficiency, intuitiveness, and rationality of the variable sets.

E. MODELING TECHNIQUE

Last chapter, the modeling techniques commonly used in financial scoring models were introduced and described. Each technique is capable of accurately discriminating between failed and nonfailed firms, employing varied methodologies and generating varied outputs. These methodologies (the underlying mechanics of the techniques) and outputs (the scale of the dependent variable) have implications for both the developer of the model and the user.

1. The Issue

What are the comparative advantages of the statistical techniques applied to the task of failure prediction? Each technique is chosen by the model developer for a variety of reasons: it may best fit the tested hypothesis, it may be a new application of the technique or a comparison study of two or more techniques, it was used in a prior relevant work, or as one author described "[it] would eventually have to be explained to the conference delegates and the authors felt it would be easier to explain discriminant analysis than logit/probit analysis..." (Keasey and Watson, 1986). The last reason is not meant to belittle the rationale used by the authors, rather it may be pertinent. For some users, the ability to explain the rationale for a decision that is based on the outcome of a model is necessary; selecting a modeling technique that is understandable is important.

There are six main techniques to be discussed: univariate discriminant analysis (UDA), multivariate discriminant analysis (MDA), conditional probability models (CP) (i.e., logit and probit regression analysis), recursive partitioning (RP), indexing, and

artificial intelligence (AI). The use of these techniques in the literature reviewed by the author is presented in Table 4, below.

Each technique has inherent assumptions and limitations which affect the development and use of the model. For those developing a model or users who are interested in an in depth discussion of the issues, there exists a substantial literature which

Model	UDA	MDA	CP	RP	AI	Index	Other
Beaver, 1966	X						
Altman, 1968		X					
Wilcox, 1971							X
Edmister, 1972		X					
Deakin, 1972	X	X					
Altman & McGough, 1974		X					
Blum, 1974		X					
Ohlson, 1980			X				
Dambolena & Khoury, 1980		X					
Mensah, 1983		X	X				
Rose & Giroux, 1984		X					
Zavgren, 1985			X				
Frydman, Altman & Kao, 1985				X			
Keasey & Watson, 1986		X					
Dopuch, Holthausen & Leftwich, 1987			X				
Moses & Liao, 1987		X				X	
Lau, 1987			X				
Barniv & Raveh, 1989							X
Seaman, Young & Baldwin, 1990		X	X				
Moses, 1990						X	
Dagel & Pepper, 1990		X					
Platt & Platt, 1990			X				
Koh & Killough, 1990		X					
Koh, 1991			X				
Goss, Whitten & Sundaraiyer, 1991			X				
Schary, 1991			X				
Aly, Barlow & Jones, 1992		X	X				
Baldwin & Glezen, 1992		X					
Coats & Fant, 1993		X			X		
Cormier, Magnan & Morard, 1995		X	X	X			
Platt, 1995			X				

Table 4. Modeling Techniques Used in Failure Prediction Models

will be referenced throughout this section. This thesis will address the following issues: advantages and disadvantages for each technique, the interpretability of the contribution of individual variables, the scale of the output, underlying statistical assumptions and their implications on model use, and the apparent usefulness of the technique for discriminating failure from nonfailure.

2. The Literature

a. Univariate Discriminant Analysis (UDA)

The use of UDA was introduced by Beaver (1966) in his seminal work on the use of financial scoring models to predict business failure. It is a useful technique when the predictive ability of a single variable is of particular interest. It does not suffer from a problem of multicollinearity commonly found in multivariate models. It is simple to employ, and easy to understand.

Its principal criticism is that the use of a single variable fails to capture the multidimensional complexity of the financial status of a business. Performing UDA on several variables may yield conflicting messages. To combat these two criticisms, the body of work has expanded to include the use of multidiscriminant analysis (MDA) and indexing. Recursive partitioning (RP) is also a relative of UDA in that it can be viewed as a hierarchy of univariate discriminations between the groups.

In depth discussions of UDA can be found in Beaver (1966), Altman (1968), Deakin (1972), Zavgren (1983), and Jones (1987).

b. Multivariate Discriminant Analysis (MDA)

Altman (1968) pioneered the use of MDA in business failure prediction, addressing some of the criticisms of Beaver's univariate approach to forecasting bankruptcy. The main benefit of MDA is that it captures the multidimensional complexity of the firm, however, it does suffer from several weaknesses, mainly due to the inherent assumptions in the technique.

The two main assumptions in MDA are that the independent variables are distributed multivariate normal, and the covariance matrices of the two groups are equivalent. The first assumption is frequently violated and always will be when using a dummy variable (0, 1) or certain financial ratios (those restricted to a value of $0 \leq x \leq 1$). Remedial actions, such as variable transformations, can be taken to minimize the problem, but may distort the message provided by the variable. The second assumption can be corrected by using quadratic discriminant analysis rather than the linear form. Failure to correct for either of these assumptions will affect tests of the significance of the model.

The accuracy of the linear form versus the quadratic form has been debated in the literature and several works have set out to test them by building models using each technique employing the same data (e.g., Rose and Giroux, 1984; Seaman, Young, and Baldwin, 1990; and Baldwin and Glezen, 1992). The results are inconclusive, implying

that the form is a matter of choice for the developer and user, with the quadratic perhaps having greater appeal due to lack of need to have equal covariance matrices.

The restriction on the usefulness of dummy variables makes researchers reluctant to include qualitative or macroeconomic information which frequently take the form of dummy variables in an MDA model.

The ability to interpret and test the coefficients assigned to the variables depends on whether the assumptions have been met. Since the assumptions are normally violated, it is difficult to assess the individual contribution of a particular variable. As Ohlson (1980) wrote, "A violation of these conditions, it could perhaps be argued, is unimportant (or simply irrelevant) if the only purpose of the model is to develop a discriminating device." Many users, however, would appreciate the ability to interpret the contributions of the individual predictor variables. The presence of multicollinearity also affects the ability to interpret the effects of the individual variables. Unfortunately, the use of financial ratios essentially ensures the presence of a high degree of multicollinearity, as individual actions of firms translate across multiple areas of the financial statements. The use of factor analysis to reduce the variable set is very helpful in this regard.

MDA has been criticized for producing an output that lacks intuitive interpretation; the z-score, without a comparative basis, is meaningless. The user or developer of the model must be cognizant of prior probabilities and the cost of errors in order to gain a meaningful interpretation of the score. The score does provide a benefit not found in some other techniques in that it is ordinally ranked: a firm more likely to fail will generate a lower score than one less likely to fail. If the user is making decisions about a firm relative to other firms, so long as the costs of errors are equal across the population, than the ordinal ranking does have value.

Despite the criticisms noted above, MDA is the most frequently used technique. The principal reason is its accuracy despite violations of the inherent assumptions. In depth discussions of MDA can be found in Altman (1968), Joy and Tollefson (1975), Eisenbeis (1977), Collins and Green (1982), Zavgren (1983), Jones (1987), and Altman, et. al. (1987).

c. Conditional Probability Models

The conditional probability model has an intuitive appeal. The model assumes the midrange probabilities are more sensitive to changes in the independent variables than are the extremes; what Collins and Green (1982) call "the 'threshold' property that the bankruptcy forecasting problem logically requires...". Looking at the cumulative probability density function, Figure 6, There exists a critical midrange region in which a given change in value will produce a large change in the probability of failure, yet in the tails, the same given change has a relatively minor effect. To illustrate, a 0.1 increase in the debt-to-equity ratio of a business with a 0.05 ratio would not suggest the onset of a solvency problem. Likewise, a 0.1 increase when the ratio is already 0.8 is not likely to be

much worse for the already heavily indebted firm. But if the debt-to-equity ratio is currently 0.4, perhaps the same 0.1 increase constitutes the breaking point for the firm.

Another intuitive appeal is the probabilistic output from the model.

Overcoming the criticism that the meaning of the MDA's z-score output is vague, the conditional probability model provides a continuous (0 to 1) probability estimate that the firm will fail during the specified time period. The technique also is not encumbered by the assumptions inherent in MDA regarding equal covariance matrices and multivariate normal independent variables.

The principal disadvantage of conditional probability models is that the curvilinear nature of the model makes variable interpretation complex. One cannot simply estimate a change in probability by multiplying the change in an independent variable by its coefficient. A partial derivative is needed to get an accurate assessment of the effect of an incremental change in the value of an independent variable.

Four works were reviewed that compared MDA models to conditional probability models. Seaman, Young, and Baldwin (1990) and Aly, Barlow and Jones (1992) found their MDA models performed slightly better than their conditional probability models. On the other hand, Mensah (1983) and Cormier, Magnan, and Morard (1995) studies were inconclusive.

In depth discussions of conditional probability models can be found in Collins and Green (1982), Zavgren (1983 and 1985), Jones (1987), and Altman, et. al. (1987).

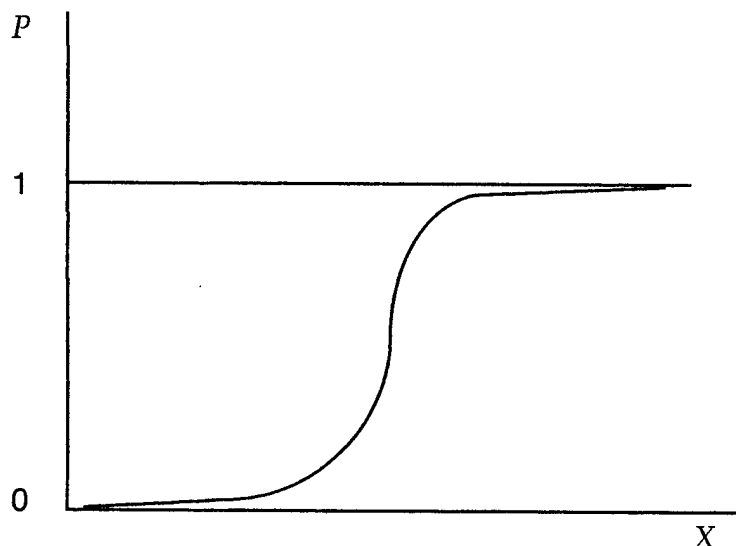


Figure 6. Cumulative probability density function

d. Recursive Partitioning (RP)

Frydman, Altman, and Kao (1985) introduced recursive partitioning to the field of failure prediction. Using financial ratios, they built two classification trees of varying complexity which performed better than discriminant models developed using the same data.

An advantage to RP is that no assumptions need to be made regarding the distributions of the variables, making it conducive to the use of financial ratios and dummy variables. The technique can also incorporate prior probabilities and costs of errors in development, rather than having to consider them in interpreting the output, as in MDA.

There are two disadvantages to RP. First is the scale of the output: RP provides neither a discriminant score nor a probability, rather the firm is simply classified as failed or nonfailed. This classification scheme is not useful if the user is comparing the relative strength of firms or wishes to obtain some gauge of the level of distress or probability of failure. Second, there is no means to evaluate the significance of the variables used to discriminate. The forward stepwise selection procedure does not permit the interpretation of relative significance based solely on the tree's hierarchy, and since it also allows a variable to reenter the classification tree, the ability to interpret its significance is confounded.

In the development of an RP model, one scenario becomes immediately apparent: the forward stepwise classification process could conceivably continue until each observation resides in a unique node. This would provide for the most accurate discrimination of the development sample, but this overfitting would compromise the generality of the model. To avoid this problem, many models must be derived of varying complexity using cross-validation procedures to reach an optimum trade-off between discriminating accuracy and stability across samples.

In depth discussions of RP can be found in Frydman, Altman, and Kao (1985), Barniv and Raveh (1989), and Cormier, Magnan, and Morard (1995).

e. Index Models

An index model is appealing in its simplicity in construction and application. It gives the relatively simple technique of UDA a multidimensional aspect without violating assumptions inherent in MDA. Zavgren (1985) noted regarding UDA, "The main difficulty with [the] approach is that classification can take place for only one ratio at a time. The potential exists for finding conflicting classifications of any given firm according to various ratios." The indexing technique – given a sufficient variable set, preferably based upon factor analysis – alleviates this criticism.

There are several advantages of indexing over MDA. First, is that extreme values in the data will affect the value of the output in MDA, whereas the index simply recognizes a cut-off score which is unaffected by extreme values. Second, the contributions of individual variables is unambiguous since they are based on separate UDA computations. Similarly, multicollinearity is not a problem; however this could be a

disadvantage in that interactions between the variables is not measured and there may be rich signals in the correlation between the variables. But these are mutually exclusive issues: MDA can take advantage of the multicollinearity to enhance discriminating ability, but loses the interpretability of individual variable contributions; indexing gains the interpretability, but loses the ability to capture the interactions between the variables. The user must decide which is most important.

In depth discussions of indexing can be found in Moses and Liao (1987) and Moses (1990).

f. Artificial Intelligence (AI)

Models developed using AI have multiple advantages and disadvantages. A key benefit of these models is their ability to process qualitative data as efficiently as quantitative data. They are also capable of learning from new data and becoming more sophisticated in their discernment. These systems are not adversely affected by extreme values, assumptions regarding probability distributions, or interrelationships between variables.

There are drawbacks, however. The field of artificial intelligence is still not widely accepted and is viewed skeptically by much of the public-at-large. Modeling problems include inconsistencies of reasoning within and between the "experts" used as templates for the programming. The selection of an expert or experts is also problematic (i.e., whose "mental model" is the *correct* one?). These systems also fail to exercise "common sense;" the enormity of human intellect is not possible to program, yet may very well come into play even in seemingly simple analyses. The systems may also be considered too smart: they do not know when they do not have sufficient information. Whereas a human would know to ask additional questions or collect another data point, the artificial intelligence system assumes it has sufficient reasoning capability. Perhaps the greatest failing is that the programming requires knowledge of the process of human intellect in order to repeat it. The fields of psychology and sociology have enormous gaps in the knowledge about knowledge.

In depth discussions of the use of artificial intelligence in financial scoring models can be found in Coats (1988) and Coats and Fant (1993).

g. Other techniques

Two other techniques were noted by the author in reviewing the literature. Wilcox (1971) used a gambler's ruin model to predict bankruptcy (see the previous discussion on cash flow theory in Section A), but found that developing the necessary probability estimates was so uncertain and unreliable, he abandoned the technique. Barniv and Raveh (1989), using the same data and independent variables Frydman, Altman, and Kao (1985) used to develop their RP model, created a new nonparametric approach to failure prediction. Unlike RP, it provided a continuous score, and also provided greater separation of the means of the groups than MDA. It was also more accurate. Despite these apparent advantages, it has not been employed elsewhere in the literature.

3. What We Know

There have been six common modeling techniques employed in the literature. The most commonly used are MDA and conditional probability models. Used in almost equal proportions, they account for 75 percent of the models reviewed by the author.

Frequently, assumptions inherent in a technique are violated, affecting interpretation of coefficients, variables, perhaps even the output of the model. Often, however, these violations do not affect the discriminating ability of the model. A caveat is presented to the user of these models to ensure, if the interpretation of the contribution of individual variables is important, that the modeling technique and the particular model's adherence to the assumptions of that technique are scrutinized.

The interpretation of individual variable's contributions is difficult in nearly all models, impossible in RP and AI. This is not an issue for a UDA model or an index. The user or developer of the model must take this point into consideration before ascribing causality or degrees of influence on the output to particular variables.

Some models provide the user with a score, others a probability, still others a mere classification. The advantage of the score and probability is that the model's output for several firms can be compared and firms ranked based on likelihood of failure. The models that yield only a classification are less useful in this regard.

Ultimately, what is important to the user and developer is the discriminating ability of the model, its accuracy. This topic will be discussed next.

F. VALIDATION

The final dimension is the quality of the models with respect to their ability to discriminate failed from nonfailed firms. The approach was discussed last chapter; in short, it entails two parts. The first is an evaluation of the fit of the model to the development data. The second is the model's performance on both the development data and a new, or simulated to be new, sample of data. There will also be a discussion of error rates and the costs of errors, as they relate to the validity of a model.

1. Fit of the Model

The Issue: how statistically significant is the model and how well does it discriminate between failed and nonfailed firms in the development sample? The quality of a model's reported results should be evaluated using relevant statistical techniques. The issue relates to how well the model captures the differences between the failed firms and nonfailed firms in the development sample. A portion of the fit was discussed last section in the analysis of the evaluation criteria for the independent variables. This section discusses the evaluation of the model as a whole.

There are two ways to test the fit of the model. The first method is the use of appropriate measures of statistical significance. These tests are analogous to the R^2 and F-statistics used in evaluating linear regression models. The second method for testing the fit

of the model is to observe how well the model discriminated the groups in the development sample.

The Literature. The appropriate technique for testing the statistical validity of the model varies, of course, with the modeling technique employed. For models developed using MDA, Jones (1987) recommends the use of the canonical correlation to "measure the percentage of the variation in discriminant scores 'explained' by the variance between the groups." He also recommends the use of Wilk's lambda statistic for a measure of overall statistical significance. The early MDA research (e.g., Altman, 1968; Edmister, 1972; and Deakin, 1972) reported Wilk's lambda converted to F- statistics. All were very significant (to at least 95%). More recent studies have reported the canonical correlation (e.g., Koh and Killough (1990) which was significant to 99.9%) or Wilk's lambda reported as "Chi-squared" statistics (e.g., Baldwin and Glezen (1992) whose models were significant to at least 95%).

Conditional probability models, logit and probit regression, should be tested for significance using the likelihood ratio test; an analogous (or pseudo) R^2 is also a useful statistic for the model's explanatory significance. Both statistics have been cited frequently by the models' developers. Ohlson (1980) reported likelihood ratios of 0.72, 0.80, and 0.84 for his three logit regression models. Zavgren (1985) reported likelihood ratios which were all significant to 99% for each prediction year of her model. Dopuch, Holthausen, and Leftwich (1987) reported the Chi-squared statistic on the log likelihood ratio significant to 99.9%; they also reported a pseudo R^2 of 0.189. Lau (1987) offered a "probabilistic prediction score" for her five state model. In recent years, the analogous R^2 has become more common. Platt and Platt (1990) cited an analogous R^2 of 0.56; Cormier, Magnan, and Morard (1995) cited a pseudo R^2 of 0.84; and Platt (1995) reported an analogous R^2 of 0.86.

The second way of measuring the fit of the model, regardless of modeling technique, is its ability to discriminate the firms in the development sample. Table 5 lists the classification accuracy of those models reviewed by the author, data is provided for both the reported accuracy of the classification of the development sample and the accuracy reported by the author for a validation test (to be discussed next subsection). Some explanation of the figures is necessary. First, the figures reported are classification accuracy rates; that is, the percentage of firms classified correctly whether they are failed or nonfailed. The cost of errors was considered to be equal for both Type I and Type II errors, so the error rates were additive. Second, in most cases, the figures reported are those cited specifically by the author. In those cases when the authors cited multiple models and formulations of models using the same data, the data cited are for the formulation and accuracy rates which are most representative of the study as a whole, or those accuracy rates most often cited in the literature. For example, Dambolena and Khoury (1980) created 7 models, some using just financial ratios, others using ratios and

Model	Development Sample					Validation Sample				
	Years Prior to Failure					Years Prior to Failure				
	1	2	3	4	5	1	2	3	4	5
Beaver, 1966	87	79	77	76	78					
Altman, 1968	95	72	48	29	36					
Edmister, 1972	93					93				
Deakin, 1972	97	95	95	79	83	78	94	88	77	85
Blum, 1974	93	88	80	86	83	95	80	70	80	69
Ohlson, 1980	96	95	92*			85	85			
Dambolena & Khoury, 1980	91	85	83	89						
Mensah, 1983	V1					V1				
Rose & Giroux, 1984 LMDA						92	88	91	94	89
Rose & Giroux, 1984 QMDA						84	87	78	74	80
Zavgren, 1985	82	83	72	73	80	69	69	69	69	69
Frydman, Altman & Kao, 1985	V2					V2				
Keasey & Watson, 1986	75									
Dopuch, Holthausen & Leftwich, 1987	V2					V2				
Moses & Liao, 1987 MDA	73					71				
Moses & Liao, 1987 Index	81					79				
Lau, 1987	96	92	90			80	79	85		
Barniv & Raveh, 1989	V2					V2				
Seaman, Young & Baldwin, 1990 LMDA						69				
Seaman, Young & Baldwin, 1990 QMDA						78				
Seaman, Young & Baldwin, 1990 CP						66				
Moses, 1990	96					96 / 100				
Dagel & Pepper, 1990	97					93				
Platt & Platt, 1990	90					87 / 90				
Koh & Killough, 1990	92					88				
Koh, 1991	99					99				
Goss, Whitten & Sundaraiyer, 1991	86									
Aly, Barlow & Jones, 1992 MDA	90	89	92							
Aly, Barlow & Jones, 1992 CP	100	100	98							
Baldwin & Glezen, 1992 †	89	87	83	80	79					
Coats & Fant, 1993 MDA						88	83	84	84	
Coats & Fant, 1993 AI						95	93	86	82	
Cormier, Magnan & Morard, 1995 MDA	77									
Cormier, Magnan & Morard, 1995 CP	76									
Cormier, Magnan & Morard, 1995 RP	75									
Platt, 1995	77					73				
V1 = varies depending on model formulation					† = measured in quarters before failure					
V2 = varies depending on cut-off scores and costs of errors					* = within two years of failure					

Table 5. Classification Accuracy Rates (in Percentages)

standard deviations; the table reports the classification accuracy for the latter model applied to four years of data; and of the 21 models developed by Blum (1974), the table reports the accuracy of the model using a four year interval of data to predict up to five years prior to failure.

In a table such as this, where models are held up to comparison against others, it is proper to point out that some of the cited models were not constructed in a manner to attain high accuracy. Rather, they were developed to compare varying modeling techniques or to test the predictive ability of certain variable transformations. Examples include Seaman, Young, and Baldwin (1990), Coats and Fant (1993), and Cormier, Magnan, and Morard (1995).

What we know. We have seen that the developers of models generally apply appropriate tests of statistical significance to their models. These tests have shown a very high degree of statistical significance implying that the models adequately capture differences between the groups of failed and nonfailed firms.

We also know that the models cited in the literature generally perform quite well in classifying the development data. Accuracy rates are particularly high one to two years prior to failure, and tends to fall off beyond the third year. Some authors have presented their accuracy results in increasingly complex ways, introducing the issues surrounding the costs of errors. This will be discussed in the next subsection where the models will be evaluated on their performance outside the development sample.

2. The Performance of the Model

The second aspect of model validation is its performance, particularly in discriminating a sample of firms different from that used in the model's development. There are two main issues to be discussed: the first issue relates to the model's actual performance discriminating a second sample of data, the second issue relates to error rates and the costs of errors. For the user, these issues are critical. The user is concerned about how well the model will perform in application to a new set of data and whether it will give economically efficient results given a cost structure for misclassifications.

a. Performance On a New Sample

The Issues: how well does the model perform on a sample distinct from the sample used in its development? The last chapter introduced the fundamental choice facing the model developer when selecting a sample on which to apply the model to assess its performance: the model can be validated on a sample taken from within the development data or from an entirely new set of data. Due to the scarcity of data on failed firms, or limitations imposed by the construct, the use of a validation sample taken from within the development sample may be necessary for purely practical reasons. There are two common options available to the researcher if the choice is made to use within sample data: the use of a split sample or the Lachenbruch technique.

If the developer has the luxury of a larger population or a longer period of time (and the construct of the research permits it), the use of a second, outside sample is the preferred method of validation. There are two common options facing the researcher: a holdout sample from the same period of time as the development sample, or a second sample from a subsequent time period.

There is a third method of model validation of importance to the user or other researchers, but does not involve choices made during development. Models are occasionally tested at a later time by the same or another researcher using data from a distinct population. These tests are done strictly as a test of the older model, to compare multiple models, or perhaps the older model is being used as a benchmark for comparison. Regardless of the reason, it is helpful to see how robust these models are when tested on different populations.

The Literature. Referring back to Table 5, we see in the second set of columns of data the results of validation testing. The testing reported here is that done by the models' authors and reported in the work that introduced the model. Multiple techniques were used to conduct these validations so comparisons should not be made hastily.

Among those using a within-sample validation technique are Rose and Giroux (1984) who used the Lachenbruch method; it is interesting to note that they reported only the validation sample accuracy, which naturally tends to be lower than the accuracy of the development sample. Platt (1995), Koh (1991), and Platt and Platt (1990) also used the Lachenbruch technique to validate their models; the latter reported 87% accuracy using the Lachenbruch technique and also reported 90% accuracy using a sample of new data. Dagel and Pepper (1990) used a split sample design, taking 25 of the 29 pairs in the development sample and retesting them.

The proper technique for testing the significance of a recursive partitioning model is cross-validation. The procedure involves breaking the sample into several subsets, recomputing the model using all but one subset of firms and reclassifying the remaining ones; this is repeated once for each subset. Frydman, Altman, and Kao (1985) reported using a five-fold cross validation technique. The other recursive partitioning model, Cormier, Magnan, and Morard (1995) did not report the use of a cross-validation technique.

The use of a hold-out sample from the development population or entirely new data has been slightly more common. Unusual in his approach, Deakin (1972) validated his model using data that *predated* the development data. It would seem more useful to use later data, as most others have done. Those employing data from a subsequent time period include Mensah (1983), Barniv and Raveh (1989), and Koh and Killough (1990). Rather than use a set of data from a later period collected along the same criteria as the development sample, as others have done, Koh and Killough applied their

model to a random sample of 400 firms, 14 of which happened to be bankrupt. Finally, those models using a holdout sample from within the time period of the development sample include Blum (1974); Dopuch, Holthausen, and Leftwich (1987); Moses and Liao (1987); Lau (1987); and Moses (1990) who used two different holdout sampling methods.

Of those models that have been tested again after validation, Altman (1968) has been tested more often than any other. As the pioneering MDA model (and despite numerous criticisms) it has taken on the role of benchmark for other model developers. Table 6 lists models that have been retested in the literature and the results of those tests. The classification accuracy figure reported for the model's author is the validation accuracy conducted by the author and reported in the original work, if applicable (i.e., the right hand column information from Table 5). The data clearly show that the models' performance generally declines significantly when applied to a population distinct from that which was used for its development and initial validation.

Several of the retests (e.g., Moses and Liao, 1987; Christensen and Godfrey, 1991; and Bowlin, 1995) used data representing DoD contractors. The others were tested with samples more closely resembling the development data, differing primarily in that they derived from a later time period.

Model <i>Tested by</i>	Years Prior to Failure				
	1	2	3	4	5
Beaver, 1966	87	79	77	74	78
<i>Deakin, 1972</i>	80	84	72	76	73
Altman, 1968	95	72	48		
<i>Altman & McGough, 1974</i>	82				
<i>Moyer, 1977</i>	75				
<i>Doukas, 1986</i>	80/92	68/84	80/73		
<i>Moses & Liao, 1987</i>	70				
<i>Christensen & Godfrey, 1991</i>	56				
<i>Bowlin, 1995</i>	56 / 63				
Zavgren, 1985	69				
<i>Christensen & Godfrey, 1991</i>	47				
<i>Bowlin, 1995</i>	19				
Moses & Liao, 1987	79				
<i>Christensen & Godfrey, 1991</i>	56				
Dagel & Pepper, 1990	93				
<i>Christensen & Godfrey, 1991</i>	50				
<i>Bowlin, 1995</i>	70				

Table 6. Classification Accuracy of Retesting

What we know. We can see from both the validation tests performed by the authors, and again when retested at a later time, that the performance of models is generally worse than the accuracy reported for the development sample. This is not surprising at all, considering the methods in which the models are developed. They are expected to fit data from a similar population better than a distinct one. What is surprising is the consistency of the accuracy of Beaver's single variable, cash flow to total debt. Overall, the amount of retesting that has occurred is discouraging; except for the DoD specific research, there has been nominal retesting done.

The techniques used for validation are commendable. More often than not, the models are retested with out of sample data, generating a better picture of the performance of the model than could be obtained with in sample data. The decision to use a completely random sample by Koh and Killough (1990) was a bold attempt at testing a model in "real world" conditions. The user of a model will be applying it with complete uncertainty as to the actual outcome, as the random sample implies.

b. The Costs of Errors

The Issue: how have model developers addressed the fact that, for different applications of the model, the costs of errors differ? Chapter III defined the types of errors (Type I and Type II), and introduced the issue of the costs of errors, operationalized in Equation 3. Recall, the equation showed that the cost of errors is the sum of the costs of each type of error times the probability of committing each type of error. Depending upon the use for the model, the costs may be dramatically different. Also affecting the model is the fact that the probability of failure is very low in a random population. These varying costs and probabilities must be considered by the developer and user of the model to ensure economically efficient results.

$$EC = (P_I * P(F) * C_I) + (P_{II} * P(NF) * C_{II}) \quad \text{Eq. 3}$$

The Literature. Tables 5 and 6 were constructed by assuming the costs of errors to be equal and simply summing the percentage of Type I and Type II errors. These assumptions are common within the literature and form the most reasonable basis for comparison of the models. For the user of a model, however, these assumptions are clearly naive. Certainly, the model is being applied for an economic reason, and the costs associated with the performance of the model will vary by type of error. Only a few researchers have given this much consideration.

Note in Table 5, the label V2 assigned to Frydman, Altman, and Kao (1985), Dopuch, Holthausen and Leftwich (1987), and Barniv and Raveh (1989), which indicated that determining an accuracy rate was dependent upon costs of errors and cut-off scores. Each of these three works presented tables which describe the relative accuracy of the model depending upon the costs assigned to each type of error. This treatment of model accuracy is much richer in information content than simply assuming an equal cost

and probability for each type. Similarly, Zavgren (1985) presented her model's accuracy in graphical form, varying the cut-off scores and costs of errors to permit the reader to determine the accuracy relevant to them. Only in deference to the conventional reporting of classification accuracy, did she compute the figures which are listed in Table 5.

Selecting the proper cut-off score for the model output in an environment of unequal costs of errors and prior probabilities depends upon the modeling technique. For models using discriminant analysis, Jones (1987) states "The cutoff becomes equal to $\ln [P_I * C_I / P_{II} * C_{II}]$ ", where the symbology is the same as in Equation 3. For CP models, the cut off for discriminating when costs and probabilities are equal is 0.50; when these change, the cut-off probability changes. For example, if the cost of a Type I error is four times the cost of a Type II error, the cutoff used would be a 1:4 ratio, or 0.20. Now, a firm scoring a probability of 0.2 would be classified as failed and the costs would be equal for each type of misclassification. (Jones, 1987) And for the models developed using RP, the costs of errors (and prior probabilities of membership in each category) can be specified in model development; the classification "tree" will vary depending upon the specifications.

What we know. We know that the issue of costs of errors and prior probabilities affect the economic performance of the models. For reasons of simplicity and comparability, the literature has generally assumed equal costs and used matched pair samples to simulate equal prior probabilities. While this technique has made the comparison of models relatively easy, it has resulted in distorted measures of accuracy and hidden their usefulness in practice.

It is encouraging to see some authors, recognizing the differences in costs and probabilities, have reported results suitable to a wider audience. It would be best if all developers recognized this issue and reported accuracy across the spectrum. Techniques that have been employed, and are useful to the reader, include bar charts (Zavgren, 1985), tables (Barniv and Raveh, 1989), and graphs (Frydman, Altman, and Kao, 1985). The user is cautioned in applying models to make adjustments to the cutoff scores to ensure economical outputs for the specific application.

G. CONCLUSIONS

The primary research question this thesis has attempted to answer is: What is the state of the art in the use of financial scoring models for the purpose of predicting business failure? The author hopefully has answered this question sufficiently through a comprehensive evaluation of the literature along six dimensions: (1) the theoretical bases for the models, (2) sampling and data collection, (3) the dependent variable and definition of failure, (4) the independent variables, (5) the modeling technique employed, and (6) the subsequent validation of the models. An examination of the state of the art has not been conducted before on as many dimensions or considering as many models.

There has been much significant research conducted in the area of failure prediction since the last examination of the state of the art (Jones, 1987). The findings of the author,

summarized below, show that the boundaries of the field have widened, that many of the unresolved issues raised by Jones have been addressed, that others still elude answer, and that new questions have been raised. Next chapter, and the thesis, will conclude with a return to the issue of failure prediction within the DoD to answer the question: What insights and implications has this examination of the state of the art provided the DoD financial analyst?

1. Theoretical Basis

There are two classes of theory currently influencing the field: theory regarding the behavior of the firm and theory regarding particular information sets. Within each of these classes are two specific theories. Behavioral theories include the cash flow theory and the events approach. Informational theories have developed taxonomies of financial ratios for describing firms and theory derived from the auditing literature regarding indicators that a firm may no longer be a viable going concern. Each of the theories most directly impacts the selection of independent variables, but has also raised issues related to the dependent variables. These theories have also provided an appreciation for factors which affect firm failure beyond the accounting issues of liquidity and the ability to meet financial obligations.

2. Sample Selection and Data Collection

Jones (1987) recommended "new efforts...should attempt to develop models for the important group of small and new firms." Unfortunately, the state of the art offers little new knowledge about this segment of the population. Considering the failure rate of smaller and newer firms, this is a particularly important group to study, unfortunately, the same characteristics which make it an important group to study impede the study: data is scarce and unreliable.

The inevitable trade-off between sample size and relevance is a continual problem for the field. The state of the art has been the acceptance of smaller samples. While the rate of failure is certainly a limiting factor, much of the research has also been plagued with problems of data availability. It is hoped that advances in information technology will alleviate the availability problem and provide for larger, while equally relevant, sample sizes. Much of the research has employed a matched pair design in the sample construction, providing the benefit of insulating the data from variables outside the realm of the intended study, but sacrificing some validity in application.

3. The Dependent Variable and Definition of Failure

Recently, the literature has moved away from traditional definitions of failure, in large part due to the application of new theory. Models have been developed predicting excessively negative shareholder returns, qualified audit opinions, debt accommodations, and other measures of financial distress. While complicating comparability of models, the

field has gained a richness in recognizing that failure is manifest in many ways besides bankruptcy.

Most of the literature has provided models designed to predict failure in a specified time frame; a minority of models have chosen to predict failure within a range of time. Each has its benefits, but the latter has more value in application as the output is less ambiguous. The range of outputs has expanded from the discrete output of the discriminant models to include probability estimates and simple dichotomous classifications. The models generating continuous outputs have been used to classify firms into two, three, and even five categories.

4. The Independent Variables

The literature still shows a strong bias toward the use of quantitative variables to comprise the information set used to predict failure. While the use of qualitative measures is still relatively sparse, there has been an increase in their use in recent years and it is expected to rise as the events approach to failure and the literature from auditing gain more acceptance. The most frequently used quantitative data are derived from the financial statements, and are normally in the form of financial ratios. The information content of those ratios usually centers around measures of profitability, liquidity, leverage, and cash flow management. The use of factor analysis to derive taxonomies of financial ratios with strong descriptive abilities has received considerable attention recently. It has been demonstrated that the financial condition of firms can be described with only a few ratios and recently they have been shown to be robust across industry segments and economic climates.

Failure, it can be argued, is more an economic event than a financial one, yet the field continues to predict its occurrence using primarily financial indicators, despite evidence that the economic climate is influential on the rate of failure. Failure is also a dynamic event so the study of changes in the financial condition of a firm seems logical; but the study of trends in the data has been done less often than one would expect. Its frequency is increasing, however. Other transformations of variables have been conducted recently to minimize the effects of some condition (e.g., size or economic state) or to enhance the predictive ability of the variable (e.g., industry-relative ratios and measures of variability).

5. The Modeling Techniques

There have been six common modeling techniques employed in the literature with the most common being multidiscriminant analysis and conditional probability models. Since Jones (1987) examination of the state of the art, the use of recursive partitioning, indexing, and artificial intelligence has emerged. The relative advantages and disadvantages of each were discussed last chapter. Frequently, assumptions inherent in a technique are violated, affecting interpretation of coefficients, variables, and even the

output of the model, but normally these violations do not affect the discriminating ability of the model. As with the scale of the output, the choice of a modeling technique is a function of the intended application.

6. Validation

The literature has shown that the developers of models generally apply appropriate tests of statistical significance to their variables and models. These tests have shown a very high degree of statistical significance implying that the models adequately capture differences between the groups of failed and nonfailed firms. The models generally perform well in classifying the development data with high accuracy rates one to two years prior to failure. Recently, accuracy results have been presented in increasingly complex ways, including consideration of the costs of errors, and more common use of out of sample data for validation. Validation tests performed by the authors, and subsequent retesting, have generally shown a decline in accuracy. The retesting that has occurred is discouraging, both in the reported accuracy of the models and the limited extent to which retesting efforts have been conducted.

H. RECOMMENDATIONS FOR FURTHER RESEARCH

The author has identified five issues which are worthy of further study. First, investigation into the events associated with failure needs to continue. This investigation should focus on the events leading up to failure, the post-mortem of failed firms, and the analysis of firms rebounding from distress, with the findings investigated for potential predictors of failure. As a business is an economic entity with self-preservation as a basic tenet, the early signs of financial distress and impending failure which may be detected by a model are also known to the management and would most certainly be met by some remedial action. Clearly, not all actions are sufficient to prevent failure, but then again, many are. It would be useful in developing a model intended to predict failure to include elements addressing this remedial action.

Second, the field has only begun to assess the predictive ability of the mental models used by auditors in assessing the going concern risk of client firms. Related to this are the models employed by bankers to determine loan default risk. Most banks do not use financial scoring models (Makeever, 1984), and models have not been conclusively shown to be superior to the bankers' judgment (Libby, 1975; Casey, 1980; Zimmer, 1980; Houghton, 1984; Chalos, 1985; and Doukas, 1986). Perhaps there are lessons to be applied to financial scoring models from the techniques used by bank loan officers.

Third, the expanding definition of failure is adding a richness to the literature and should be continued, but the cost is a loss of comparability between models. There needs to be sufficient research done on each definition to fully assess what the significant predictive variables are for each definition.

The fourth issue regards the study of small, private firms. Data availability has been a considerable problem and a solution is not apparent. But this category of business is a compelling group to study as they are the most likely to fail.

The fifth issue was alluded to in the conclusions, above. The use of more retesting of models in distinct populations, simulating the models' use in practical application, would be helpful to those who intend to employ the models. While a purely academic approach is valuable, and necessary, more of a practical approach is becoming necessary.

V. IMPROVING DEPARTMENT OF DEFENSE FINANCIAL ANALYSIS

Thus far, this thesis has introduced the current practice of financial analysis within DoD (Chapter II) and evaluated the academic literature creating a snapshot of the state of the art in the use of financial scoring models for the purpose of predicting business failure (Chapter IV). This chapter will extract from the state of the art the research that has specific relevance to DoD and, using the broader state of the art as a backdrop, reach some conclusions on how current DoD practices should be revised. In other words, the goal of this chapter is to improve current practices by taking advantage of the relevant elements of the state of the art external to DoD and incorporating the lessons learned from the body of DoD specific research.

This chapter is constructed as follows. First, the state of current practices will be reviewed. Second, the body of DoD specific research will be extracted and summarized from the larger body of academic work presented in Chapter IV. Third is a discussion on how to build better models for application by DoD commands performing financial analysis. Finally, recommendations for further research and action will be presented.

A. REVIEW OF CURRENT DOD PRACTICES

Recall from Chapter II that there are five DoD activities performing financial analysis. Two joint commands, the Defense Contract Management Command (DCMC) and the Defense Contract Audit Agency (DCAA), conduct analyses in support of the contract award and administration processes. Three service-specific commands, the Naval Center for Cost Analysis (NCCA), the Army Center for Resource Analysis and Business Practices (ACRABP), and the Air Force Office of Economic and Business Management (OEBM), perform financial analyses to support acquisition milestone reviews and on an ad hoc basis to assess the financial health of the services' respective industrial base.

Recall also that each command takes a unique approach, employing one or more of the following techniques: financial scoring models, ratio analysis, capital market information, and commercial credit scoring agencies. The level of sophistication in the analysis ranges from exclusive use of commercial credit scoring agencies to comprehensive analyses comprised of financial scoring models, ratio analysis, cash flow forecasting, and various qualitative measures. Of the financial scoring models available, only two are actively used in DoD, the Altman (1968) model and the Dagel and Pepper (1990) model.

For a detailed discussion of financial analysis in DoD, refer to Borah (1995).

B. DEFENSE INDUSTRY SPECIFIC RESEARCH

There exists in the literature a few works which compare and evaluate financial scoring models within the DoD context. Other defense related works have selected their

data from a sample comprised of or representing DoD contractors. Still others have evaluated some aspect of the financial condition of DoD contractors. This extraction of the broader literature, what could be referred to as the DoD literature, is summarized below.

1. DoD Specific Models

There have been several models developed specifically for use in DoD or with DoD contractors as the development sample. These models are presented below in chronological order.

a. Matthews (1983)

Description. Matthews (1983) examined the usefulness of the qualitative information found in annual financial reports for predicting failure of defense contractors. He did not rely on any of the theoretical bases outlined in this thesis for either the construct or the selection of variables for the model. Failure was defined as bankruptcy and a matched pair sample was compiled of 20 failed and 20 nonfailed publicly traded government contractors from various industries from the period 1970 through 1982. The model used only one independent variable, the "integrative complexity" of the firms presidents' messages contained in the annual reports. Integrative complexity derives from the behavioral sciences and asserts that those whose language exhibits high levels of complexity, normally are more complex thinkers and will consider more alternatives when problem solving. Matthews tested to see if higher integrative complexity on the part of the corporations' presidents would result in a better survival rate.

The presidents' messages were scored for their integrative complexity for five years prior to failure. There were two hypotheses tested. The first tested for a statistical difference between the failed and nonfailed firms, the second tested for a general decline in complexity as failure approached.

Findings. The first hypothesis was supported by the data: there was a statistical difference between the categories of firms. The second hypothesis was rejected: the change in complexity from year to year was not statistically significant. He concluded that "while the complexity of language in the presidents' cover letters for failing firms does not decrease as bankruptcy approaches, the integrative complexity scores for failing firms are consistently lower than those of the non-failing firms."

Key issues. Matthews commented that his sample may have been too small. He was not concerned so much with the number of firms, but that the time horizon prior to failure may have been too short (five years). The expectation of finding some homogeneity in the complexity of the language in early years and a simplification for failed firms as failure approached was not realized. He suggested that this divergence may have occurred years earlier.

b. Moses and Liao (1987)

Description. Moses and Liao (1987) developed a financial scoring model by identifying the most relevant financial dimensions for a sample of DoD contractors,

determining representative measures for those dimensions, and combining them into a failure prediction index. This was an early attempt to use a theoretical basis for the selection of specific measures for the independent variables.

Their sample consisted of a matched pair of 26 failed and 26 nonfailed small, privately held government contractors. Failed firms were defined as having filed for bankruptcy. Financial data was collected from documents supplied to government contracting agencies. Factor analysis was performed on 21 financial ratios derived from this data and which represented those frequently found significant in earlier studies; four distinct factors were identified. Various ratios representing those factors were analyzed for their discriminating and predicting ability using univariate discriminant analysis.

The resultant model was an index comprised of three ratios representing three of the four factors. A cutoff value was assigned to each ratio (derived from the univariate analysis conducted on each), and a score of one was assigned if a firm's ratio was above the cutoff, zero if it was below. A firm scoring a total of two or three was considered healthy, otherwise it was considered to be facing impending failure.

Findings. The three ratios found to be most significant at predicting failure were (1) net worth to assets, (2) working capital to assets, and (3) sales to assets. The model correctly classified 81% of the development sample and 79% of a holdout validation sample. These results were superior to those achieved using an alternative multivariate discriminant analysis technique on ratios representing the same factors.

Key issues. The sample was specifically chosen to represent small, government contractors from varying industries in a wide range of geographical areas. Studying small firms is often difficult due to data availability problems. In this case, data was obtained from three independent sources – all government agencies – all having acquired the data from the subject firms. Being unaudited, data obtained from the firms themselves is often of questionable reliability and comparability. This issue was not specifically addressed.

c. Dagel and Pepper (1990)

Description. Dagel and Pepper (1990) built a failure prediction model using a DoD specific sample. The model was developed without regard to a particular theory for the construct or independent variable selection. Employing a matched pair design, they used a sample of 29 failed (defined as bankruptcy) and nonfailed publicly traded firms.

Financial ratios were used as the independent variables. The set of potential variables was reduced from 18 ratios (selected due to their popularity in the literature) down to six using stepwise regression, judgment, and tests of the statistical significance of the variables.

The model was developed using the MDA technique. Performance was assessed in two ways. First, the discriminating ability one year prior to failure was measured. Second, the predictive accuracy of the model was determined by applying data

dating from two to five years prior to failure for 25 of the 29 firms and then assessing the historical predictive accuracy of the model over those years.

Findings. Four of the six variables used in the model were measures of liquidity. The other two variables measured the level of debt (total debt to total assets) and scale of operations (net sales to total assets). The authors recorded 97% accuracy with their development sample one year prior to failure. When tested on a historical basis (four prior years worth of data for 25 of the 29 firms), accuracy fell to 64%, 60%, 44%, and 24% for years two to five prior to failure, respectively.

Key issues. Data availability was a problem in the development of the model. Only ten of the 29 firms comprising the sample were actual DoD contractors. In order to obtain a sufficiently large sample, the remaining 19 firms were selected based upon their similarity in all other respects.

The validation technique was also unusual in that the model's predictive ability was tested by using historical data from the development sample and projecting into the present rather than using a holdout sample or projecting into the future.

d. Christensen and Godfrey (1991)

Description. Christensen and Godfrey (1991) made their most significant contribution in their retesting of DoD specific models, which will be discussed below. They also developed their own models using the same defense contractor data that was used for retesting. That data was obtained from the files of the Air Force Accounting and Finance Center for government contractors who had filed for bankruptcy. This data was augmented with additional data taken from the files of the Securities and Exchange Commission, the Wall Street Journal Index, and the Compustat database.

The independent variables were selected on the basis of theory: the ratio taxonomies suggested by Pinches, Mingo, and Carruthers (1973). They originally considered 20 ratios representing the seven categories suggested by the theory. Using both the MDA and logit regression (CP) techniques, the authors considered over 200 different formulations of models before settling on final models. The models were validated using the Lachenbruch procedure.

Findings. For both techniques, the model which was most statistically significant employed only one independent variable, cash to total assets, and was accurate to 78%. Validation accuracy was determined to be 75%.

Key issues. There were three key issues. First, of the 150 government contractors identified through Air Force files as having declared bankruptcy, sufficient data existed for only five of them, generating the need to compile additional data. The second issue is that, despite applying a theoretical basis to variable selection, the resultant models only contained one variable each. They suggested using other variables, such as capital market information and macroeconomic indicators. Third, they suggested an examination of the prior probabilities of failure and the costs of errors for government applications, particularly as they would relate to a more relevant definition of failure.

2. Other Relevant Studies

Recall from last chapter's discussion of theories regarding the information content of independent variables, Moses' (1995) study of the financial ratios of defense firms. Using factor analysis to identify patterns inherent in the ratios of a sample of defense firms, he found that there are eight basic dimensions of financial condition for these firms and he isolated financial ratios representative of those dimensions. The eight dimensions, and their representative ratios, can be classified into two broad categories as shown in Table 7.

Data was selected which covered both expansionary and recessionary times for the industry and firms were selected which crossed various industry segments. The ratios were tested and found to be stable over time, over different economic conditions, and across different industry segments.

Intensity or Success of Operations		Financial Position	
Dimension	Ratio	Dimension	Ratio
Turnover	Capital Turnover	Cash Position	Cash to Total Assets
Profitability	Return on Capital	Inventory	Inventory to Current Assets
Cash Flow	Cash Flow to Total Assets	Asset Composition	Current to Total Assets
		Liquidity	Current Ratio
		Leverage	Long-Term Debt Ratio

Table 7. Dimensions of Financial Condition of DoD Contractors

3. Research Evaluating the DoD Specific Models

There have been three recent works which have evaluated the use of financial scoring models in the DoD context. They have tested models employed by DoD with new data, they have discussed the relative merits of using financial scoring models as part of financial analysis, and they have reached differing conclusions. They are presented in chronological order. In the succeeding section, the recommendations and findings of these three studies will be incorporated into an original discussion of the issues which need to be addressed to develop and apply models in a DoD context.

a. Christensen & Godfrey (1991)

The authors looked at two models specifically developed for use in the defense industry, Dagel and Pepper (1990) and Moses and Liao (1987), compared them with two other models popular in the literature, Altman (1968) and Zavgren (1985), and then created their own based specifically on defense contractor data. (Their original models were described above.)

Using data representing defense contractors, they compiled a matched pair sample of 18 failed (defined as bankruptcy) and 18 nonfailed firms. Each of the above

models was applied to the new data and the discriminating ability of the models was measured. The results showed that the models were much less accurate in discriminating the new data than reported in the original study. Accuracy was reported last chapter in Table 6, but ranged from 47% to 56%. They attributed the relatively poor performance to oversampling of bankrupt firms in the models' developmental samples, nonstationary collinear ratios, and unequal misclassification costs¹. They recommended that to improve a models' accuracy in a DoD context, there needed to be a clearer definition of failure, the use of non-accounting data as independent variables (e.g., capital market information), and a larger sample size.

b. Bower & Garber (1994)

Under contract with the Air Force, the authors studied the problem of using models to predict failure among defense contractors. After reviewing the literature, they recommended against using statistical approaches that use historical financial data, citing impediments such as the extent and irreversibility of the current drawdown, the restructuring within the defense industry, and the small number of bankruptcies among publicly traded defense industry firms. They also state that "...studies have not been able to focus on firms in a single industry," but do not cite as references the DoD specific works noted above (e.g., Moses and Liao, 1987; Dagel and Pepper, 1990; Christensen and Godfrey, 1991).

In supporting their position against adapting currently existing models to defense use, they cited what they consider to be several unique features of the defense industry: sensitivity to the business cycle; use of progress payments; the existence of government-owned, contractor-operated assets; and extreme discrepancies between book and market values. They make three recommendations. First, in lieu of using financial scoring models, they advocate the use of financial market data, such as firm market value and private bond ratings services. Second, if models are used, they should predict financial distress rather than failure. Third, they advocate models be used for ranking firms rather than categorizing them individually.

c. Bowlin (1995)

In an approach similar to Christensen & Godfrey (1991), Bowlin compared the models currently in use in DoD against other models. Those selected were the Altman five-variable model, Altman four-variable model, Dagel and Pepper, Zavgren, and a simple bank credit risk model (an index based upon twelve financial ratios and their values relative to the industry).

Bowlin applied all five models to a sample of defense industry firms from the period 1986-1990. Of note is that none of the firms in the sample failed during this

¹ While costs do not inherently affect accuracy, the authors raise the point that when the costs of errors are unequal, the cutoff score discriminating between failed and nonfailed should be adjusted. When they made such an adjustment, the accuracy rates naturally declined as the cutoff score is no longer set to minimize errors, it is set to minimize costs.

period. Hence, only Type II errors (erroneously predicting failure when a firm is healthy) were examined. His findings are also presented in Table 6 of Chapter IV. He concluded that the best performing model was the bank credit model followed by the Altman four-variable model (a variation of the five-variable model discussed in this thesis) and the Dage and Pepper model. By relaxing the definition of failure from bankruptcy to merely deteriorating financial health, he retested the models but the results were not significantly different. He did not discourage their use in current DoD applications, but cautioned that they should not be used indiscriminantly, but rather as part of a more comprehensive analysis.

C. CONSTRUCTING A BETTER DOD FINANCIAL SCORING MODEL

Last chapter, the state of the art in the academic literature was explored, and the first two sections of this chapter have presented the DoD literature. What follows in this section is an evaluation of the construction and application of financial scoring models for predicting failure within a DoD context. The issues unique to the DoD context will be explored and the DoD literature evaluated against the backdrop of the academic literature. The result is a framework for improving financial analysis within DoD. For consistency, this evaluation will use the six dimensions developed in Chapter III and used in Chapter IV.

1. Theoretical Basis

The academic literature discussed two categories of theory and two specific theories within each category. How can these be applied in a DoD context? In the DoD literature, the only one of the four theories to be applied is the one regarding taxonomies of financial ratios. This theoretical basis has been used in model construction by Moses and Liao (1987) and Christensen and Godfrey (1991). The theory was further refined, specific to DoD, by Moses (1995). It would appear that future model development for DoD would be well served by applying the eight dimensions of financial condition identified by Moses (1995) to the selection of independent variables. (There are also benefits to be derived in the selection of a sample when using these dimensions, noted in section 2 below.)

What of the other three theories? The cash flow theory has been the most widely applied in the literature, perhaps it can offer some insight in a DoD context. Cash flow for a major defense contractor is affected by progress payments, contract payment schemes (e.g., "cost plus profit" or "fixed fee" types), government-owned and contractor-operated equipment, and the ability to sell the technology or product in other markets. With these unique characteristics, perhaps a unique approach to the cash flow theory is warranted.

John and John (1992) and John (1993) both used the cash flow theory in their models. They showed that firms in specialized industries are particularly vulnerable to financial distress; their proxies for specialized industries were the level of advertising expenses and research and development expenses. Considering the high amounts of

research and development spending among DoD contractors, the theory suggests they may be particularly vulnerable.

The events approach may be useful in a DoD context as well, particularly as it relates to the dependent variable. Many of the events associated with or preceding bankruptcy may be relevant definitions of failure in and of themselves. The auditing literature should likewise be considered. Perhaps surveys of auditors of defense contractors to assess the measures they use in forming a going concern opinion can provide insight into useful predictive measures for a model.

2. Sample Selection and Data Collection

Sampling plagues DoD specific research perhaps even more so than it does the field as a whole. The bounds of firm size, industry, and economic conditions cause the tension between sample size and relevance. The failure rate of defense contractors is very small, limiting samples, but recent findings in the literature suggest that this may be manageable. First, Moses (1995) showed that the dimensions of financial health are stable across time, macroeconomic climates, and industry segments. This suggests that samples can be broadened across these dimensions. While the dimensions are stable, the actual values seem not to be², but there may be a way to account for the differences. For example, measures of liquidity are stable across economic climates, but the actual numerical value for the measure of liquidity that best discriminates failed from nonfailed firms may differ. It may be possible to add to the model an economic indicator to account for the change in the critical values at different times.

Second, despite an apparent lack of success in the Bowlin (1995) study, there is still a way to expand the data set by relaxing the definition of failure. When the definition of failure is narrowly defined, as in bankruptcy prediction models, the sample size is restricted to firms meeting that narrow definition. But if the definition was relaxed to include a state of financial distress less severe than bankruptcy, then the sample size will increase to include those firms that were distressed, but managed to avoid bankruptcy. The appropriateness of such a relaxation was discussed above. In Bowlin's case, the definition was relaxed, but the models were not recalibrated or reformulated. Perhaps the best predictors of bankruptcy are not the best predictors of a relaxed definition of failure and that is the cause of the lackluster performance. A change in the dependent variable should be accompanied by a reevaluation of the entire model.

There still exists the issue of data availability for small firms, however. Small firms are a particularly problematic segment of the population and they receive tens of billions of dollars of business from the defense department each year. When considering small firms for large contracts, there should be a requirement for the firms to provide several years of

² Studies such as Lev (1969), Lev and Sunder (1979), Mensah (1984) and Rose, Andrews, and Giroux (1984) suggest that in differing industry segments and macroeconomic conditions, the signals provided by financial ratios will differ.

financial data for an evaluation. Over the long run, if this data were stored in a central database, it would quickly grow sufficiently large to provide a resource for further research.

Considering the changes to the structure of the defense industry recently, obtaining comparable data is more difficult. Extending the bounds of the sample along the dimension of time, for example, the financial data for Martin-Marietta is now consolidated with that of Lockheed and soon with that of Loral. The task of extracting data which meets the quality criteria described last chapter may be a considerable undertaking.

3. The Dependent Variable and Definition of Failure

The next issue requiring further study is the dependent variable: what is the best operational definition of failure for a model in a DoD context? Future research should consider the economic costs of various definitions of failure, or states of financial health, and develop models for predicting these events. The use of models such as Altman's, designed to predict bankruptcy, may be less valuable when the event of bankruptcy is moot because the greater cost to the government is associated with some other event that occurs en route to bankruptcy and should have been predicted earlier.

In essence, there are broader constructs of interest to DoD users than those which have been used in financial scoring models. For contracting applications, the operational definition may be financial distress sufficient to cause the need to make advance payments to the contractor or distress sufficient to cause delays in production due to a lack of liquidity. Similarly, for assessing the health of the industry or a member firm of the industry, a bankruptcy model may be appropriate, or a model indicating the likelihood of mergers or acquisitions, or perhaps a model assessing the employment of capital assets normally devoted to defense use.

Depending upon the application for the model, the scale of the output is a consideration. The models currently in use have been developed using MDA and provide for an output that can be ordinaly ranked, allowing for the comparison of one firm's score with another's or one firm's score at different points in time. The ability to do this is particularly important when using the model to assist in the award of a contract. Other applications may be well served by a simple dichotomous failed/nonfailed output as may be provided by a recursive partitioning or artificial intelligence model. One can imagine setting the cutoff at an appropriate level so that the model will yield a warning when a firm engaged in a long-term contract deteriorates to a certain point. To date, all of the DoD literature has provided models that yield a numerical output. Future research should be conducted on developing models that yield probability estimates or dichotomous outputs.

4. Independent Variables

If the definition of failure is changed, how does that impact the selection of independent variables? One must question whether a predictor of bankruptcy is also a

predictor of contract default, for example. The academic literature shows a broadening of the definition of failure which is useful to DoD, but these broader definitions are not yet fully tested, nor adequately relevant to DoD applications.

In constructing a model for DoD using a new definition of failure, careful attention must be paid to the selection of independent variables to capture the essence of the defined failure event. The attention should be placed on all the issues raised in the last two chapters: the information set, the selection of specific measures, and the evaluation of those measures.

Thus far, all of the models developed using DoD contractors in the sample have employed financial ratios as independent variables, with the exception of the Matthews (1983) study. While financial ratios have been most popular in the academic literature and have yielded high performing models, there may be relevant information contained in other types of variables. The DoD literature tends to support this notion. Christensen and Godfrey (1991) recommend the use of non-accounting data such as capital market information and macroeconomic indicators. Bower and Garber (1994) also advocate the use of financial market data such as bond ratings and bond yields; they also suggest investigating measures of the interdependencies between defense contractors. Another source of potential variables includes the administering contracting officers (ACOs) who work closely with the larger defense contractors – they may be able to provide DoD specific insight into signs of firm weakness. Perhaps contract related variables may be relevant which suggest contract risk rather than firm risk, e.g., contract type, contract value as a percentage of total firm revenue, subcontractor firm risk, and trends in the ratio of man-days per unit of contract value.

Any attempt to use theory other than that already applied to the DoD context may result in the use of independent variables beyond financial ratios, particularly those theories that suggest the use of more qualitative variables (i.e., the events approach and the theory derived from the field of auditing). Even in the absence of theory to suggest their use, the literature has shown several qualitative variables to be strongly associated with failure.

There has been increasing use of transformations of variables in the academic literature to enhance the predictive ability of the model. That is, the models have capitalized on the variability and trends inherent in the variables and used these features as additional independent variables. The DoD specific works have seldom incorporated such transformations.

5. Modeling Technique

The most common modeling techniques in the academic literature are MDA and CP. In the DoD literature, all models were developed using MDA, except one. What benefit is gained by the DoD user with each technique? Or, perhaps a better way to examine the issue is to address the requirements of the DoD user and see which technique is best suited to those requirements.

In the section on the dependent variable, the notion of the scale of the output was discussed and how the scale should differ for differing applications. When comparing two firms or one firm at various points in time, it is helpful to use a technique that provides a continuous score rather than a simple dichotomous classification. Those techniques that provide a continuous score are UDA, MDA, indexing, and CP. On the other hand, RP and AI provided merely a classification of failed/nonfailed and are less useful when comparing firms.

Another consideration of the DoD user is how understandable and defensible the model is. That is, if a firm is not selected for the award of a contract or suffers some other economic loss due to a decision based in whole or in part on the output of a financial scoring model, the quality of the model may be challenged. Therefore, the model itself, its output, variables, coefficients, and cutoff scores should all be interpretable, understandable, and rational, making them defensible. In order to meet these criteria, the statistical assumptions of the modeling technique must be adhered to. In this case, MDA becomes problematic as its assumptions are normally violated. In addition, the individual coefficients of an MDA model are not interpretable, only their ratios are. Interpretability of the coefficients is also difficult with CP models due to the curvilinear nature of the probability density function, the marginal contribution of the individual variables is not a linear function. RP and AI suffer from the fact that the modeling technique drives the variable selection, allowing variables to reenter the model, confounding variable interpretation and possibly the interpretation of the cutoff scores. The dichotomous classification output is very simplistic and depends upon the sensitivity prescribed by the model developer. UDA and indexing have a distinct advantage over the other methods: each variable and coefficient is interpretable and their marginal contributions are individually measurable. The UDA and index models are simple to use and understand.

6. Validation

Of course, the final test of a model is its ability to discriminate failed firms from nonfailed firms. No model is perfectly capable in this regard and all perform worse as the model is applied to firms different than those used for the model's development. The question at issue is: how good is good enough? As discussed in part B above, there has been some retesting of models (both general ones and ones developed using DoD firms as their development sample) using new samples of DoD contractors. The results of the retesting have demonstrated a marked decline in performance over that reported in the original research. In fact, those models used in DoD or developed using a DoD sample have performed with 50 to 70 percent accuracy in retesting.

One could argue that 50 to 70 percent is not much better than a chance classification and the financial scoring model is of questionable value over other forms of financial analysis. This argument may be valid if the analyst is merely applying the model without supplemental analysis and taking its output at face value. It has been noted in the literature

and will be repeated here, the use of financial scoring models should be a part of a larger assessment of the financial health of a firm and should not be relied upon solely. Furthermore, for a device intended to predict a future event, 50 to 70 percent accuracy in tests conducted under "real world" conditions is fairly significant. A poorly constructed model would be expected to perform much worse than 50 percent accuracy.

Previously, the merits of a model that provides for a continuous output that can be ordinarily ranked was discussed. When these types of models are applied, so long as they are more accurate than chance, they will most likely provide a reasonable basis for claiming that one firm is healthier than another.

The academic literature on validation and retesting is not fully developed and is less so when examining a specific industry application such as DoD. Any decision to embrace or abandon the use of financial scoring models in a DoD context is premature. Further testing of existing models and rigorous validation of future models are necessary before any such judgment can be made.

D. CONCLUDING REMARKS

Both the academic literature and the DoD literature have been evaluated and the state of the art has been presented. The field of failure prediction through the use of financial scoring models has progressed significantly since the literature was last critically evaluated (Jones, 1987), yet the application of recent findings in the literature to a DoD context has not occurred.

On the other hand, there has been considerable attention paid in recent years to critical analysis of the use of financial scoring models in a DoD context. Rather than building new, more capable models, the literature has evaluated the models currently in use. This is a helpful exercise and has provided insight into both the performance of the models and their applicability in differing circumstances.

The next step is to incorporate the suggestions made above:

- The definition of failure must be critically reevaluated in a DoD context. Those DoD personnel benefiting from the application of financial scoring models (e.g., contracting officers, financial specialists, policy makers) should work with model developers to define failure and to suggest variables that are predictive of that definition.
- The use of financial ratios is well accepted and the recent findings on the dimensions of financial condition within the defense industry should be incorporated in future model development.
- Models, their variables, coefficients, outcomes, and cutoff scores should be defensible.

- The model should be rigorously validated and retested or recalibrated as necessary as conditions within the DoD context change.
- There should also be a coordinated effort for collection of data on small businesses to compile an accurate and complete database, not just for failure prediction use, but for countless other potential research efforts concerning small firms in business with the government.

LIST OF REFERENCES

- Altman, E. I., "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol. XXIII, No. 4, Sept. 1968, pp. 589-609.
- Altman, E. I. and T. P. McGough, "Evaluation of a Company as a Going Concern," *The Journal of Accountancy*, Dec. 1974, pp. 50-57.
- Altman, E. I., R. B. Avery, R. A. Eisenbeis, and J. F. Sinkey, Jr., *Application of Classification Techniques in Business, Banking and Finance*, Jai Press, Greenwich, 1981.
- Altman, E. I., "Accounting Implications of Failure Prediction Models," *Journal of Accounting, Auditing, and Finance*, Vol. 6, No. 1, Fall 1982, pp. 4-19.
- Aly, I. M., H.A. Barlow, and R. W. Jones, "The Usefulness of SFAS No. 82 (Current Cost) Information in Discriminating Business Failure: An Empirical Study," *Journal of Accounting, Auditing, and Finance*, Vol. 7, No. 2, Spring 1992, pp. 217-229.
- Asquith, P., R. Gertner, and D. Scharfstein, "Anatomy of Financial Distress: An Examination of Junk-Bond Issuers," *The Quarterly Journal of Economics*, Vol. CIX, No. 3, August 1994, pp. 625-658.
- Baldwin, J. and G. W. Glezen, "Bankruptcy Prediction Using Quarterly Financial Statement Data," *Journal of Accounting, Auditing, and Finance*, Vol. 7, No. 3, Summer 1992, pp. 269-285.
- Beaver, W. H., "Financial Ratios as Predictors of Failure," *Empirical Research in Accounting: Selected Studies, supplement to Journal of Accounting Research*, 1966, pp. 71-111.
- Blum, M., "Failing Company Discriminant Analysis," *Journal of Accounting Research*, Spring 1974, pp. 1-25.
- Borah, David C., "Financial Analysis of Private Sector Firms Within the DoD," Master's Thesis, Naval Postgraduate School, June 1995.
- Bower, A. G. and S. Garber, *Statistical Forecasting Of Bankruptcy of Defense Contractors -- Problems and Prospects*, The Rand Corporation, Santa Monica, 1994.
- Bowlin, W. F., "An Empirical Comparison of Models for Identifying Fiscal Stress in the Defense Industry," *Journal of Cost Analysis*, Spring 1995, pp. 105-129.
- Bulow, J. I. and J. B. Shoven, "The Bankruptcy Decision," *Bell Journal of Economics*, Vol. 9, No. 2, Autumn 1978, pp. 437-456.
- Campisi, S. and K. T. Trotman, "Auditor Consensus in Going Concern Judgments," *Accounting and Business Research*, Vol. 15, No. 60, Autumn, 1985, pp. 303-310.
- Casey, C. J., "the Usefulness of Accounting Ratios for Subjects' Predictions of Corporate Failure: Replication and Extensions," *Journal of Accounting Research*, Autumn 1980, Vol. 18, No. 2, pp. 603-613.

- Chalos, P., "Financial Distress: A Comparative Study of Individual, Model, and Committee Assessments," *Journal of Accounting Research*, Autumn 1985, pp. 527-543.
- Chen, K. H. and T. A. Shimerda, "An Empirical Analysis of Useful Financial Ratios," *Financial Management*, Spring 1981, pp. 51-60.
- Christensen, D. and I. Godfrey, "Predicting Financial Distress in Defense Contractors," *Proceedings of the Acquisition Research Symposium*, 1991.
- Coats, P. K., "Why Expert Systems Fail," *Financial Management*, Autumn 1988, pp. 77-85.
- Coats, P. K. and L. F. Fant, "Recognizing Financial Distress Patterns Using a Neural Network Tool," *Financial Management*, Autumn 1993, pp. 142-155.
- Collins, R. A. and R. D. Green, "Statistical Methods for Bankruptcy Forecasting," *Journal of Economics and Business*, March 1982, pp. 179-190.
- Cormier, D., M. Magnan, and B. Morard, "The Auditor's Consideration of the Going Concern Assumption: A Diagnostic Model," *Journal of Accounting, Auditing, and Finance*, Vol. 10, No. 2, Spring 1995, pp. 201-222.
- Dagel, H. W. and R. M. Pepper, "A Financial Distress Model for DOD Hardware Contractors," *Journal of Parametrics*, Vol. 10, No. 1, Feb. 1990, pp. 5-40.
- Dambolena, Ismael A. and S. J. Khoury, "Ratio Stability and Corporate Failure," *Journal of Finance*, Sept. 1980, pp. 1017-1026.
- Deakin, E. B., "A Discriminant Analysis of Predictors of Business Failure," *Journal of Accounting Research*, Spring 1972, pp. 167-179.
- DeAngelo, H. and L. DeAngelo, "Dividend Policy and Financial Distress: An Empirical Investigation of Troubled NYSE Firms," *Journal of Finance*, Vol. 45, 1990, pp. 1415-1432.
- Dickerson, O. D., and M. Kawaja, "The Failure Rates of Business," published in *The Financing of Small Business: A Current Assessment*, I. Pfeffer, ed., Macmillan Company, 1967.
- Dopuch, N., R. W. Holthausen, and R. W. Leftwich, "Predicting Audit Qualifications with Financial and Market Variables," *The Accounting Review*, Vol. LXII, No. 3, July 1987, pp. 431-454.
- Doukas, J., "Bankers versus Bankruptcy Prediction Models: An Empirical Investigation, 1979-82," *Applied Economics*, Vol. 18, 1986, pp. 479-493.
- Edmister, R.O., "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction," *Journal of Financial and Quantitative Analysis*, March 1972, pp. 1477-1493.
- Eisenbeis, R. A., "Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics," *The Journal of Finance*, Vol. XXXII, No. 3, June 1977, pp. 875-900.

Furtado, E. P. H. and V. Karan, "Causes, Consequences, and Shareholder Wealth Effects of Management Turnover: A Review of the Empirical Evidence," *Financial Management*, Summer 1990, pp. 60-75.

Gilson, S. C., K. John, and L. Lang, "Troubled Debt Restructurings: An Empirical Study of Private Reorganization of Firms in Default," *Journal of Financial Economics*, Oct. 1990, pp. 315-353.

Giroux, G. A. and C. E. Wiggins, "An Events Approach to Corporate Bankruptcy," *Journal of Bank Research*, Vol. 15, No. 3, August 1984, pp. 179-187.

Goss, E., A. Whitten, and V. Sundaraiyer, "Forecasting with Lotus Based Logit Regression Models," *The Journal of Business Forecasting*, Spring 1991, pp. 19-22.

Hawkins, David F., *Corporate Financial Reporting and Analysis*, 3rd ed., Irwin, Boston, 1986.

Houghton, K. A., "Accounting Data and the Prediction of Business Failure: The Setting of Priors and the Age of Data," *Journal of Accounting Research*, Spring 1984, pp. 361-368.

Houghton, K. A. and R. Sengupta, "The Effect of Prior Probability Disclosure and Information Set Construction on Bankers' Ability to Predict Failure," *Journal of Accounting Research*, Autumn 1984, pp. 768-775.

Hudson, John, "An Analysis of Company Liquidations," *Applied Economics*, Vol. 18, 1986, pp. 219-235.

John, K., L. H. P. Lang, and J. Netter, "The Voluntary Restructuring of Large Firms in Response to Performance Decline," *The Journal of Finance*, Vol. XLVII, No. 3, July 1992, pp. 891-917.

John, K. and T. A. John, "Coping With Financial Distress: A Survey of Recent Literature in Corporate Finance," *Financial Markets, Institutions, and Instruments*, Dec. 1992, pp. 63-78.

John, T. A., "Accounting Measures of Corporate Liquidity, Leverage, and Costs of Financial Distress," *Financial Management*, Autumn 1993, pp. 91-100.

Johnson, C. G., "Ratio Analysis and the Prediction of Firm Failure," *The Journal of Finance*, Dec. 1970, pp. 1166-1168.

Jones, F. L., "Current Techniques in Bankruptcy Prediction," *Journal of Accounting Literature*, Vol. 6, 1987, pp. 131-164.

Keasey, K. and R. Watson, "The Prediction of Small Company Failure: Some Behavioural Evidence for the UK," *Accounting and Business Research*, Vol. 17, No. 65, Winter 1986, pp. 49-57.

Khanna, N. and A. B. Poulsen, "Managers of Financially Distressed Firms: Villains or Scapegoats?," *The Journal of Finance*, Vol. L, No. 3, July 1995, pp. 919-940.

- Kitfield, J., "Industrial Base Gets Smaller, Colder," *Government Executive*, August 1995, pp. 39-47.
- Koh, H. C. and L. N. Killough, "The Use of Multiple Discriminant Analysis in the Assessment of the Going-Concern Status of an Audit Client," *Journal of Business Finance & Accounting*, Vol. 17, No. 2, Spring 1990, pp. 179-192.
- Koh, H. C., "Model Predictions and Auditor Assessments of Going Concern Status," *Accounting and Business Research*, Vol. 21, No. 84, Autumn 1991, pp. 331-338.
- Lau, A., "A Five-State Financial Distress Prediction Model," *Journal of Accounting Research*, Vol. 25, No. 1, Spring 1987, pp. 127-138.
- Lev, B., "Industry Averages as Targets for Financial Ratios," *Journal of Accounting Research*, Autumn 1969, pp. 290-299.
- Lev, B. and S. Sunder, "Methodological Issues in the Use of Financial Ratios," *Journal of Accounting and Economics*, Vol. 1, Winter 1979, pp. 187-210.
- Libby, R., "Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence," *Journal of Accounting Research*, Spring 1975, pp. 150-161.
- Makeever, D. A., "Predicting Business Failures," *Journal of Commercial Bank Lending*, Vol. 66, 1984, pp. 14-18.
- Matthews, W. M., "Predicting Business Failure: Identifying High-Risk Contractors," Master's Thesis, Naval Postgraduate School, June 1983.
- Menon, K. and K. B. Schwartz, "An Empirical Investigation of Audit Qualification Decisions in the Presence of Going Concern Uncertainties," *Contemporary Accounting Research*, Spring 1987, pp. 302-315.
- Mensah, Y. M., "The Differential Bankruptcy Predictive Ability of Specific Price Level Adjustments: Some Empirical Evidence," *Accounting Review*, Vol. LVIII, No. 2, April 1983, pp. 228-246.
- Mensah, Y. M., "An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study," *Journal of Accounting Research*, Vol. 22, No. 1, Spring 1984, pp. 380-395.
- Mintz, J., "Military Trims Ranks, Budget by Outsourcing," *San Jose Mercury News*, December 24, 1995, pp. 1E-2E.
- Moses, O. D., "On Bankruptcy Indicators from Analysts' Earnings Forecasts," *Journal of Accounting, Auditing, and Finance*, Vol. 5, No. 3, summer 1990, pp. 379-404.
- Moses, O. D. and S. S. Liao, "On Developing Models for Failure Prediction," *Journal of Commercial Bank Lending*, March 1987, pp. 27-38.
- Moses, O. D., "Basic Dimensions of Financial Condition Within the Defense Industry," Naval Postgraduate School Technical Report No. NPS-SM-95-008, Monterey, CA, Nov. 1995.

Moyer, R. C., "Forecasting Financial Failure: A Re-examination," *Financial Management*, Spring 1977, pp. 11-17.

Ofek, E., "Capital Structure and Firm Response to Poor Performance," *Journal of Financial Economics*, Vol. 34, 1993, pp. 3-30.

Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol. 18, No.1, Spring 1980, pp. 109-131.

Opler, T. C. and S. Titman, "Financial Distress and Corporate Performance," *The Journal of Finance*, Vol. XLIX, No. 3, July 1994, pp. 1015-1040.

Pinches, C. E., K. A. Mingo, and J. K. Caruthers, "The Stability of Financial Patterns in Industrial Organizations," *The Journal of Finance*, May 1973, pp. 389-396.

Platt, H. D., *Why Companies Fail*, D.C. Heath & Company, 1985.

Platt, H. D. and M. B. Platt, "Development of a Class of Stable Predictive Variables," *Journal of Business Finance and Accounting*, Vol. 15, Spring 1990, pp. 31-53.

Platt, H. D. and M. B. Platt, "A Note on the Use of Industry-Relative Ratios in Bankruptcy Prediction," *Journal of Banking and Finance*, Vol. 15, 1991, pp. 1183-1194.

Platt, H. D., "A Note on Identifying Likely IPO Bankruptcies: A Symphonic Paradox," *Journal of Accounting, Auditing, and Finance*, Vol. 10, No. 1, Winter 1995, pp. 71-79.

Rose, P. S., W. T. Andrews, and G. A. Giroux, "Predicting Business Failure: A Macroeconomic Perspective," *Journal of Accounting, Auditing, and Finance*, Vol. 6, No. 1, Fall 1982, pp. 20-31.

Rose, P. S. and G. A. Giroux, "Predicting Corporate Bankruptcy: An Analytical and Empirical Evaluation," *Review of Business and Economic Research*, Vol. 19, No. 2, Spring 1984, pp. 1-12.

Schary, M. A., "The Probability of Exit," *RAND Journal of Economics*, Vol. 22, No. 3, Autumn 1991, pp. 339-353.

Schwartz, K. B., "Accounting Changes by Corporations Facing Possible Insolvency," *Journal of Accounting, Auditing, and Finance*, Vol. 6, No. 1, Fall 1982, pp. 32-43.

Scott, J., "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models," *Journal of Banking and Finance*, No. 5, 1981, pp. 317-344.

Seaman, S. L., D. M. Young, and J. N. Baldwin, "How to Predict Bankruptcy," *The Journal of Business Forecasting*, Fall 1990, pp. 23-28.

Senge, P., *The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday/Currency, New York, 1990.

Van Frederikslust, R. A. I., *Predictability of Corporate Failure*, Leiden, Boston, 1978.

Weintraub, J., "Expert Systems in Government Administration," *AI Magazine*, Vol. 10, No. 1, Spring 1989, pp. 69-71.

Wilcox, J., "A Gambler's Ruin Prediction of Business Failure Using Accounting Data," *Sloan Management Review*, Spring 1971, pp. 1-10.

Wilcox, J., Book Review of "Corporate Bankruptcy in America" by E. I. Altman, *The Journal of Accountancy*, Feb. 1973, pp. 92-93.

Wruck, K. H., "Financial Distress, Reorganization, and Organizational Efficiency," *Journal of Financial Economics*, Vol. 27, 1990, pp. 419-444.

Zavgren, C. V., "The Prediction of Corporate Failure: The State of the Art," *Journal of Accounting Literature*, Vol 2, 1983, pp. 1-38.

Zavgren, C. V., "Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis," *Journal of Business Finance & Accounting*, Vol. 12, No. 1, Spring 1985, pp. 19-45.

Zavgren, C. V. and G. E. Friedman, "Are Bankruptcy Prediction Models Worthwhile? An Application in Securities Analysis," *Management International Review*, Vol. 28, 1988, pp. 34-44.

Zimmer, I., "A Lens Study of the Prediction of Corporate Failure by Bank Loan Officers," *Journal of Accounting Research*, Autumn 1980, pp. 629-636.

Zmijewski, M. E., "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, Vol. 22, Supplement 1984, pp. 59-86.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
8725 John J. Kingman Rd., STE 0944
Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
Naval Postgraduate School
411 Dyer Rd.
Monterey, CA 93943-5101

3. Professor O. Douglas Moses 1
Code SM/MO
Naval Postgraduate School
Monterey, CA 93943-5101

4. Professor Shu S. Liao 1
Code SM/LC
Naval Postgraduate School
Monterey, CA 93943-5101

5. LCDR Philip J. Candreva, SC, USN 1
607 Overlook Court
Wexford, PA 15090